First Principle Study of the Electronic Structure of Semiconductors for Photovoltaic Applications: Organic-Inorganic Perovskites

by

Brendan Morningstar

A thesis submitted in partial fulfillment
of the requirements for the degree
of Master of Science (MSc) in Physics

The Faculty of Graduate Studies
Laurentian University
Sudbury, Ontario, Canada

THESIS DEFENCE COMMITTEE/COMITÉ DE SOUTENANCE DE THÈSE
Laurentian Université/Université Laurentienne
Faculty of Graduate Studies/Faculté des études supérieures

Title of Thesis Titre de la thèse	First Principle Study of the Electronic Structure of Semiconductors for Photovoltaic Applications:Organic-Inorganic Perovskites
Name of Candidate Nom du candidat	Morningstar, Brendan Derek
Degree Diplôme	Master of Science
Department/Program Département/Programme	MSc Physics
Date of Defence Date de la soutenance	August 16, 2018

APPROVED/APPROUVÉ

Thesis Examiners/Examinateurs de thèse:

Dr. Mohamed Azzouz
(Supervisor/Directeur(trice) de thèse)

Dr. Rizwan Haq
(Committee member/Membre du comité)

Dr. Ubi Wichoski
(Committee member/Membre du comité)

Dr. Abdelilah Benyoussef
(External Examiner/Examinateur externe)

Approved for the Faculty of Graduate Studies
Approuvé pour la Faculté des études supérieures
Dr. David Lesbarrères
Monsieur David Lesbarrères
Dean, Faculty of Graduate Studies
Doyen, Faculté des études supérieures

ACCESSIBILITY CLAUSE AND PERMISSION TO USE

Acknowledgement

First and foremost, I would like to convey my individual regard and gratefulness to Professor Mohamed Azzouz with whom I have been deeply inspired. Pursuing a career in this field was largely motivated by his immense wisdom and transparent insight into the physical concepts. I feel overly fortunate to have had a kind, considerate, and infinitely patient professor throughout my graduate studies.

I would like to express my earnest joy and praise to my brother, Mason. To Brent and Mitchell, I express my thanks in the name of the many *drafts* (and occasional struggles) it took to finish this work.

I would like to thank my parents, Tricia and Kevin, for loving and supporting me to their fullest. In their nourishment and guidance, I have enjoyed the ultimate freedom both academically and personally. I would also like to express my heartfelt thanks to Sue-Ellen and Terry for their constant support and encouragement.

I would like to express my very special thanks, affection, and admiration to Thomas. Her excellent support and guidance during this work deserves the highest praise.

Finally, I would like to express my sincere and warmest gratitude to Sarah for her faith and helping me improve every aspect of my life. I want to thank her for being with me in every happy and difficult time and for sharing the most charming moments of her life. For these reasons and more, this thesis is dedicated to you, my dear.

Abstract

The ill effects of climate change affect all trends, and the steps taken in the drive to reduce global emissions will reverberate for thousands of years. It is among the most significant and urgent problems we face, and so it is immensely important to call upon existing and near future technologies for generating clean electricity. For now, the most talked-about renewable energy source is solar. It is a massive resource by any standard and it has the potential to play an essential role in decreasing the dependency on crude oil and reducing fossil fuel emissions. Today, the best-performing perovskite cell has reached a power conversion efficiency of 22.1%. This unprecedented rise in efficiency for a photovoltaic technology suggests a sunny outlook, but before a large-scale deployment of the technology, there are still some real questions that must be addressed. The best performing perovskite cells contain lead, which is very toxic and damaging to the environment, and are unstable in humid conditions. Also, the fundamental working of these materials is still largely unknown. The technological base of photovoltaics is becoming progressively dependent on complicated materials, and so it is important to systematically investigate the nature of the electronic structure. In the present work, the electronic structure of five perovskite compounds, $MAPbBr_3$, $CsPbX_3$ (X=Cl, Br, I) and $RbPbI_3$, are systematically studied from first principles using the all-electron, full potential, linearized augmented plane wave ((L)APW) + local orbitals (lo) method as implemented in the WIEN2k code. It is noted that: (i) the band gap of ABX3 increases when A changes from MA to Cs; (ii) as X changes from Br to Cl to I, the band gap increases; and (iii) as A changes from Cs to Rb, the band gap mostly remains the same.

Contents

List of Figures

List of Tables

Chapter 1

Introduction and Overview

Our world is sustaining itself today on the combustion of fossil fuels. In 2014, 66% of the world's electricity generation was produced by coal, oil, and natural gas [1]. Fossil fuels are however finite - expend them for long enough and global resources will eventually dry up. Concerns surrounding this risk have persisted for many years. Imagine what would happen if the oil supply were suddenly exhausted. There would be no more gasoline for cars, or heating in many houses, or fertilizers to grow our food. There would be an immediate collapse of the world's economy and in a matter of a few months humanity would be taking back to the middle-ages. Of course there is a high probability that this will not happen rapidly but over a period of a few (5-10) decades [2]. While this is disheartening and a pressing issue, there is another crucial limit to fossil fuel production: climate change.

Since the industrial revolution, over 400 billion metric tonnes of carbon dioxide have been thrown out into the atmosphere [3]. This is significantly altering its chemical composition and severely affecting the climate of the planet [4]. Efforts to conserve energy, though highly commendable and even essential for advanced countries, have been more than offset by population increases in the less-developed world. Consequently, the worldwide energy demand and consumption continues to grow [1], and by 2030 it is expected to double [5]. This growth is alarming; with oil and natural gas reserves soon to be depleted, we are left with coal, the most abundant and polluting energy source, as the "preferred" medium to generate electricity. Figure (1.1) reflects this trend.

We have but one fragile planet to live on and it is, therefore, urgent to move to renewable energy sources. Today, the discussion is centered on the

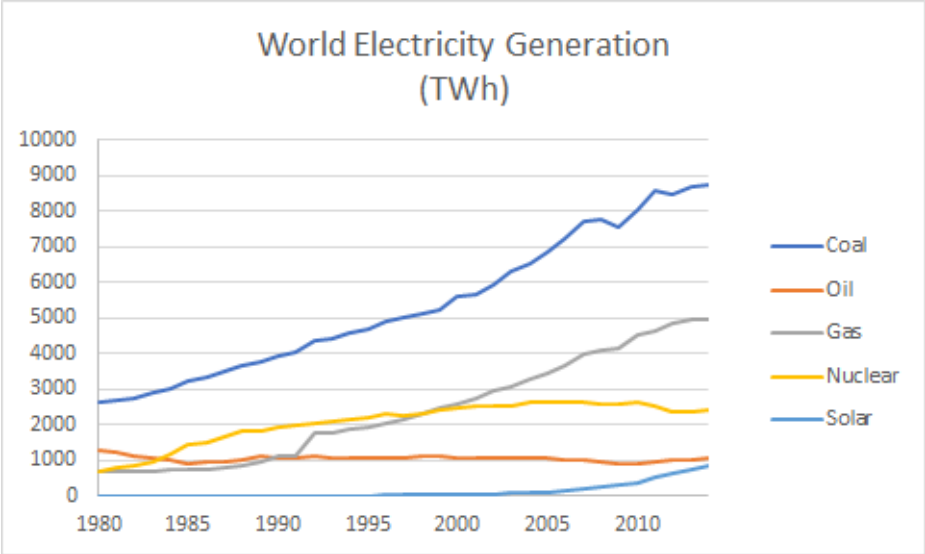


Figure 1.1: The worldwide electricity generation in billion kilowatt-hours [1].

wind, water, and solar energy [6]. On the basis of a range of options, it would be hard not to include a (quick) discussion about nuclear power - a practical and timely option for the mitigation of climate change.

The most relevant factors surrounding a large-scale nuclear presence, in short, are the public and political perception, and the potential magnitude of the economic impact. Unfortunately, these factors are viewed with rejective wariness. *First*, the precarious mixing of science and politics have somehow given the *anti*-climate change movement political resonance [7], and as a result, a gross fear of radiation, radioactive waste, and nuclear accidents, have been (wrongly) perpetuated. *Second*, nuclear plants are a lot costlier to construct than conventional fossil-fueled plants of equal generating capacity [8]. While the industry can hope that these higher capital costs will be offset by lower average fuel cost over the life of the plant, the bottom line is clear - nuclear energy is too expensive, and therefore, a non-competitive option.

The nuclear conversation is very controversial, and perhaps it has no proper place in this work. But to this image, we argue: (1) Nuclear power produces almost no carbon dioxide; (2) It is safe; (3) It is capable of supplying the energy required by a rapidly increasing population and their enhanced expectations; (3) It consumes only minimal amounts of uranium which, unlike oil and gas, is abundant everywhere in the earth's crust; and (4) It produces

just small amounts of waste. The warning here is that if not more is done to motivate the "decision-makers", nuclear power will diminish as a practical, timely, and safe option for the global energy system. And indeed, in 2014 only 11% of the world's electricity was generated by nuclear power [1].

For now, the most talked-about renewable energy source is solar. It has the potential to play an essential role in decreasing the dependency on crude oil and reducing fossil fuel emissions. About two-thirds of emissions (excluding household and industrial use of fossil fuels) are from electricity generation, transportation, and heating [9]. We already know how to generate electricity from solar energy with low emissions, and we know how to use it to provide ground transportation and can convert it to heat to provide warmth. The solar resource is massive by any standard and can easily provide the necessary energy services to accommodate the economic growth in energy consumption. A conceivable way to reduce global emission would, then, be to significantly increase the use of solar energy to generate electricity and depend more on it for transportation and heating.

The ill effects of climate change affect all trends, and the steps taken in the drive to reduce global emissions will reverberate for thousands of years. It is among the most significant and urgent problems to face, and so it is immensely important to call upon existing and near future technologies for generating clean electricity. We do not pretend to know how to foresee which perspectives might have functional importance and to what extent, but converting the underlying energy infrastructure of the world will demand tremendous efforts. This work is, therefore, motivated by the enormous potential and importance attached to photovoltaic devices to reduce global emissions and provide global energy for all purposes.

Chapter 2

Solar Energy

The sun is a potent energy source, and its radiation is by far the most significant source of energy received by the earth. Solar energy, however, includes not only direct sunlight but also several indirect forms. The hydrological cycle, for instance, depends on the evaporation of water and its subsequent return to the high ground as rain; eventually drawn out by gravity, returning to lakes, oceans, streams, rivers, and ponds. Photosynthetic fuels and the energy derived from the wind are also solar energy fixed because their energy ultimately derives from the sun: plant matter depends on the transformation of solar energy through photosynthesis; wind energy depends on the heating of land, air, and water. With the exception of nuclear, tidal, and geothermal, solar energy in one form or another is the source of all energy on this planet.

Until the late nineteenth century, when coal displaced the biomass fuels, the world depended on renewable solar energy in its many forms. Water-powered machinery lined the waterways to provide mechanical strength for such purposes as sawing wood and grinding grain; sunlight and wood, a source of light in the shadows, provided warmth and comfort. Beast of burden offered public transportation and mechanical power; the wind enabled ships to carry passengers and freight. It was a very different world than our present one, but few environmental problems are credited to that time. It should come as no surprise, then, that there is renewed interest in solar energy.

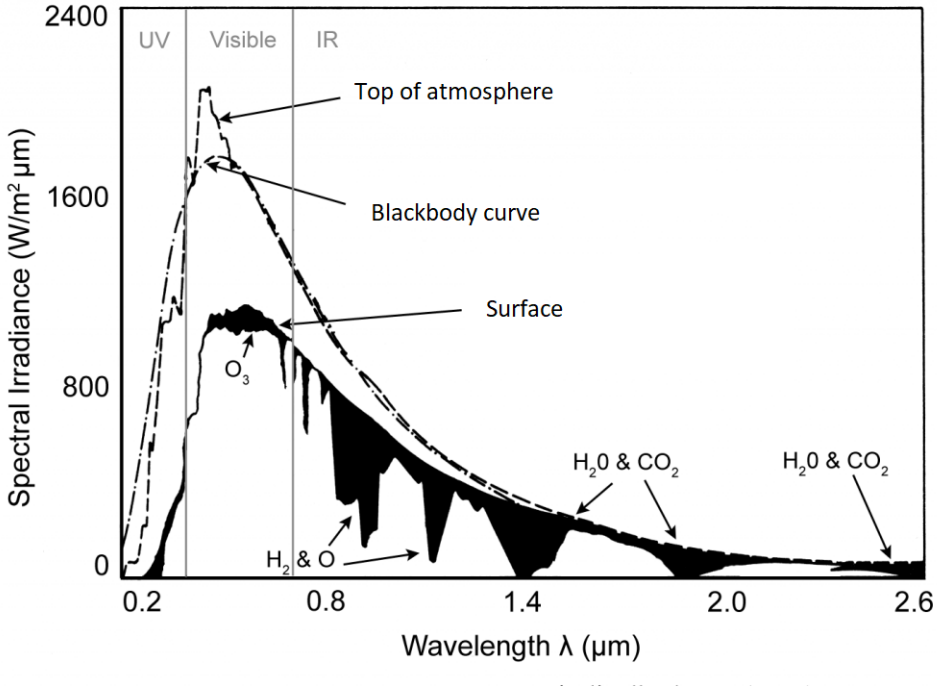


Figure 2.1: The solar spectrum at the top of the atmosphere, the theoretical blackbody curve, and the solar spectrum at the earth's surface. Absorption and scattering regions are indicated in black [10].

2.1 The Solar Resource

The sun is an engine of thermonuclear fusion and has been shining for nearly 5 billion years at the center of our solar system. It generates immense quantities of energy which, in principle, catalyzes the chemical reactions required for life. Today, however, the sun comes up, and in just one hour the world receives enough energy to power its current needs for a year. The solar resource is massive, and what is more, it costs nothing and is available to at least some extent everywhere in the world.

The surface of the sun is a rolling sea of plasma instability, called the photosphere, and approximates a blackbody at about 6000 K [11]. The temperature of the surface determines the spectral distribution of solar radiation, which is commonly reported as a function of the wavelength. About 10 percent is in the ultraviolet spectrum ($< 0.4\mu$m), 40 percent is in the visible

spectrum ($0.4 - 0.7\mu$m), and 50 percent is in the infrared ($> 0.7\mu$m); see Figure (2.1). The earth is 150×10^6km from the sun, so the overwhelming majority of this radiation - more than 99.99 percent - is not intercepted, and thus continues to travel into space beyond. At the top of the earth's atmosphere, the average intensity of the solar radiation is 1366Wm^{-2}. This quantity is the so-called solar constant, and varies (by about 3 percent) over the course of the year, from perihelion in early January to aphelion in early July.

The sunlight that actually strikes the *surface* of the earth, however, has been both scattered and absorbed by atmospheric molecules to different degrees and at various wavelengths. These interactions give rise to sharp dips in the spectrum, depicted in black in Figure (2.1), and the overall intensity is significantly reduced. The air mass factor quantifies these attenuation effects. When the sun is directly overhead, the radiation that reaches the surface corresponds to an air mass factor of 1 (AM1). As the sun moves across in the sky, creating deeper shadows, the light passes through a greater thickness of atmosphere and loses more energy; the standard representation of midday light is AM1.5, and it corresponds to about 1000Wm^{-2}. Indeed, the condition in which no light passes through the atmosphere - reported as the solar constant - corresponds to an air mass of 0 (AM0).

The solar resource is widely distributed over the earth's surface. Figure (2.2) shows a map of the yearly averaged solar radiation reaching the surface of the globe. Across the world's land and ocean surfaces, there is 6500 TW of solar energy available to power photovoltaics; however, the practical deliverable solar power is only about 340 TW.

2.2 Photovoltaic Devices

Solar photovoltaics (PV) are the most widely distributed sun-based electric technology in the world today. Fueled by light energy to produce electricity, solar cells operate near ambient temperature with no moving parts. The pristine elegance of the technology carries with it another key advantage: It is unaffected by scale. A utility-sized installation is no less efficient per unit area than a residential system. This contrasts with other means of electricity generation, such as fossil-fueled plants, which lose efficiency at a reduced scale. While the most common solar cell technologies today are based on crystalline silicon, the development of new alternative methods with better

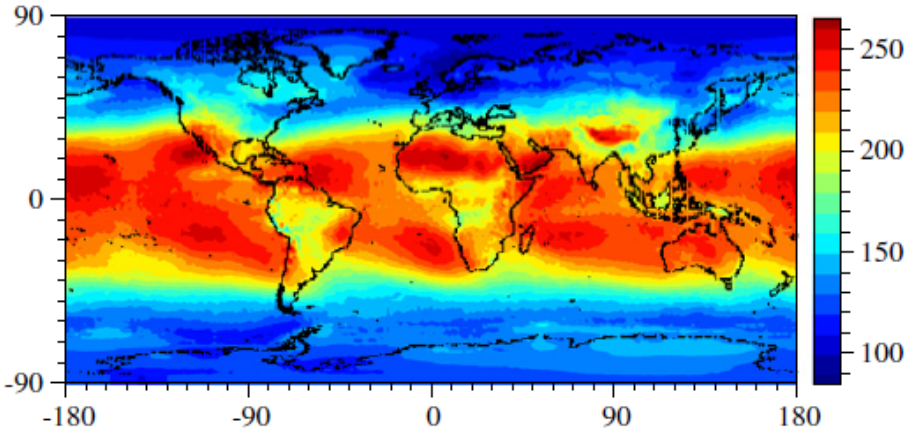


Figure 2.2: The global map of the yearly averaged downward surface solar radiation (Wm^{-2}) [6].

cost targets, lower material use, and improved efficiency, is in active research. In this section, we review the development of the different formations of photovoltaic devices.

2.2.1 First-Generation: Crystalline Silicon

Electrons in a crystal are arranged in energy bands that may be separated by definite energy gaps in which no electron orbital exists. Such forbidden regions are called band gaps and are of such essential importance to the theory of the electronic properties of solids that even elementary knowledge of them can help us understand the differences between conductors and insulators. In Chapter 4, we will go into the theory of electronic structure at some length, and we shall see why these gaps occur. For now, suppose that the number of states in an energy band is known (which is just the number of allowed values of the wavevector $\mathbf{k}$ in the first Brillouin zone).

A full band with an energy gap above it is non-conducting. That is not to say a net current cannot flow; by applying an electric field, some electrons could be excited into states representing a net current. However, with the band gap, a finite excitation energy is required to carry electrons into the upper band, and this, in general, cannot be supplied by small electric fields. In this case, the solid is an *insulator*. If on the other hand, the energy gap

is small, thermal excitations (in the energy range of the order of k_BT) may carry a small number of electrons into the upper band. The solid would have observable, temperature dependent, electrical conductivity, and so the material is a *semiconductor*. A partially filled band (where there is only one electron per atom to be placed into two spin states per unit cell) can easily carry current. In this case, the conductivity does not depend much on temperature, and the material is a common *metal*.

The first generation solar cell is based on Si , mainly single semiconductor crystals. This structure is derived from the face centered cubic (FCC) lattice by attaching a basis of two atoms at the positions (0,0,0) and (0.25,0.25,0.25) relative to the lattice points, as seen Figure (2.3). With four electrons per atom, there are eight electrons per unit cell, and four *bands* to fill. A quantitative account of these bands can be fixed from a number of theoretical perspectives (see Chapter 4), however, the qualitative prospect is simple and well known in the theory of chemical bonds.

It is known that when the Si atoms are brought together, their 3s- and 3p-orbitals - the free atom - combine to give four orbital wavefunctions which tend to concentrate the electron density in the region between neighboring ions, forming covalent bonds. These electrons, however, are not truly localized in the bonds. Like free electrons in a metal, their wavefunctions extend throughout the entire crystal, but with a highly concentrated electron density (equivalent to a pair of electrons) in the neighboring regions between ions. This means that there is a small, but non-zero, probability that the electrons can be located in another band: the conduction band. This conduction band is separated from the four valence bands by an energy gap of the order of 1 eV.

Density of States

In a semiconductor, the *density of states* - the number of electrons per unit energy range - has a big gap, and a finite excitation energy ($\approx$1 eV) is required for an electron to break free of its bound state to participate in conduction. Here, the range of energy is split into two - from the top of the valence band ϵ_v, to the bottom of the conduction band ϵ_c and onwards. Knowledge of the density of states and the band gap is intimately related to the understanding of solar cell efficiency; in fact, the quantitative representation of this scheme is fundamental to all considerations of the transport properties of semiconductors. Fortunately, this calculation is simple enough.

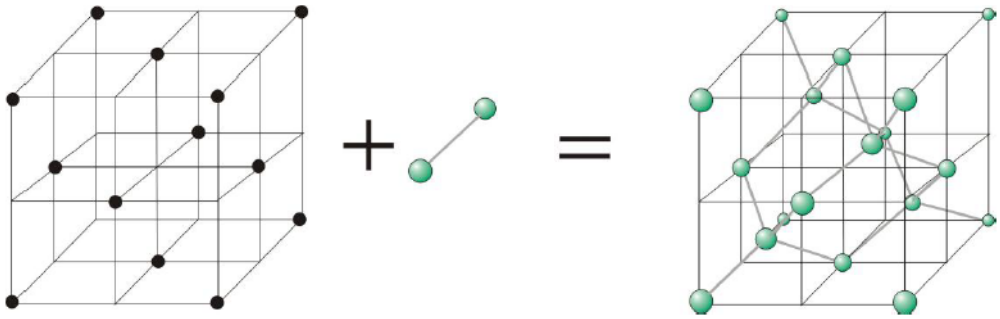


Figure 2.3: The crystal structure of silicon consists of the FCC lattice with a basis of two silicon atoms, forming the diamond structure.

In the one-electron approximation, the system has the energy levels

$$\epsilon = \sum_{\mathbf{k}} f_{\mathbf{k}} \epsilon_{\mathbf{k}}, \tag{2.1}$$

where $f_{\mathbf{k}}$ is the number of electrons in the state $\mathbf{k}$ (with spin s), and energy $\epsilon_{\mathbf{k}}$. If there are, say nV, total electrons in the system, then it must be that

$$\sum_{\mathbf{k}} f_{\mathbf{k}} = nV. \tag{2.2}$$

We state here without proof that for a system in thermodynamic equilibrium, the average number of electrons in the state $\mathbf{k}$ is given by the Fermi-Dirac distribution function [12]:

$$f^0(\epsilon_{\mathbf{k}}) = \frac{1}{e^{(\epsilon_{\mathbf{k}}-\mu)/\mathbf{k}_{\mathrm{B}}T} + 1}. \tag{2.3}$$

It follows, then, that the total density of electrons per unit volume of crystal can be written as

$$n = \int f^0(\epsilon) N(\epsilon) d\epsilon, \tag{2.4}$$

where the integral replaces the summation in Equation (2.2) for the simple reason that $\mathbf{k}$ goes over nearly a continuum of values. Since there is an energy gap in semiconductors, the "band gap" condition equivalent to Equation (2.4) is given by

$$\int_0^{\epsilon_v} N_v(\epsilon)d\epsilon = \int_0^{\epsilon_v} f^0(\epsilon)N_v(\epsilon)d\epsilon + \int_{\epsilon_c}^{\infty} f^0(\epsilon)N_c(\epsilon)d\epsilon, \tag{2.5}$$

where $N_v(\epsilon)$ and $N_c(\epsilon)$ are the density of states in the valence band and conduction band respectively. Notice that the range of integration of energy does not include states in the energy range of width ϵ_{gap}.

When an electron is excited into the conduction band, it leaves behind a space called a *hole*. Accordingly, Equation (2.5), which can be written as [12]

$$\int_0^{\epsilon_v} \big[1 - f^0(\epsilon)\big]N_v(\epsilon)d\epsilon = \int_{\epsilon_c}^{\infty} f^0(\epsilon)N_c(\epsilon)d\epsilon \tag{2.6}$$

simply states that the number of electrons excited into the conduction band is equal to the number of holes left in the valence band;

$$n_h = n_e. \tag{2.7}$$

To push this idea just a bit further, consider the bracket and the integrand term, $f^0(\epsilon)$, in (2.6). For all practical purposes, they can be written as [12]:

$$1 - f^0(\epsilon) \approx \exp\big[(\epsilon - \mu)/\mathbf{k}T\big] = \exp\big[-(\epsilon_c + \mu_\nu)/\mathbf{k}T\big] \tag{2.8}$$

and

$$f^0(\epsilon) \approx \exp\big[-(\epsilon - \mu)/\mathbf{k}T\big] = \exp\big[-(\epsilon_c + \mu_c)/\mathbf{k}T\big], \tag{2.9}$$

where μ_ν is the location of the Fermi level below the bottom of the conduction band, and μ_c is the location of the Fermi level above valence band.

With this, Equation (2.6) is somewhat simplified, and it follows that

$$n_h = \int_0^{\epsilon_\nu} [1 - f^0(\epsilon)]N_\nu(\epsilon)d\epsilon = e^{-\mu_\nu/\mathbf{k}T} \int_0^{\infty} N_\nu(\epsilon)e^{-\epsilon_\nu/\mathbf{k}T}d\epsilon_\nu \tag{2.10}$$

and

$$n_e = \int_{\epsilon_c}^{\infty} f^0(\epsilon)N_c(\epsilon)d\epsilon = e^{-\mu_c/\mathbf{k}T} \int_0^{\infty} N_c(\epsilon)e^{-\epsilon_c/\mathbf{k}T}d\epsilon_c. \tag{2.11}$$

These Equations will be valid, so long as $\epsilon_{gap} \gg \mathbf{k}T$, for an *intrinsic* semiconductor where thermal excitations create the charge carriers.

Electrons and Holes

Both the electron and hole evolve in time under the influence of an applied electric field and participates in conduction. The effect of an electric and magnetic field on the hole is given by the semi-classical equations of motion [12],

$$\dot{\mathbf{r}} = \mathbf{v}(\mathbf{k}) \tag{2.12a}$$

$$\hbar\dot{\mathbf{k}} = -e\left(\mathbf{E} + \frac{1}{c}\mathbf{v} \times \mathbf{H}\right), \tag{2.12b}$$

however, the hole has an opposite dynamical behavior to that of the electron. Accordingly, since electrons are negatively charged, it follows that a hole - being a lack of an electron - must be positively charged.

The concentration of these electrons and holes is called the intrinsic carrier concentration and is indicated by n_i. It is merely the number of electrons in the conduction band, or conversely, the number of holes in the valence band. The carrier concentration depends on the temperature and the band gap of the material. Indeed, a broader gap makes it more difficult to excite carriers into the conduction band thermally, resulting in low carrier concentration and low conductivity. Alternatively, increasing the temperature creates more electron-hole pairs and increases the conductivity of the material.

The accepted measured value of n_i for *silicon* at 300 K is $9.65 \times 10^9 \text{cm}^{-3}$ [13]. While 300 K is customarily pertinent to carrier concentration in general, photovoltaics are measured at 298.15 K (25 ^0C). Therefore, a qualitative representation of Si as a function of temperature may be more relevant here [14]:

$$n_i(T) = 5.29 \times 10^{19}(T/300)^{2.54}\exp(-6726/T). \tag{2.13}$$

To say the least, Equation (2.13) demonstrates the increase in n_i with increasing temperature.

Most semiconductors are not pure and are in a state where they have been *doped* by the addition of impurities so that the balance of electrons to holes becomes stirred. Equation (2.7) is no longer true, and the statistical theory becomes more complicated [12]. Substitutionally introducing electrons to a crystal, with impunity to the lattice structure, produces a *n-type* semiconductor material if no additional holes are created in the valence band. In this

case, the majority of carriers are *negatively* charged electrons. There is, however, a flipside to this picture, and it results when a hole is left in the valence band upon doping. Accordingly, specimens with one less valence electron are called *p-type* material - where the majority of carriers are *positively* charged holes.

For instance, doping a silicon crystal with an atom of phosphorous - which can be done without harm to the bond structure - adds electrons to the conduction band. In this case, electrons are the *majority* carriers, having the highest number. The other carriers, the holes, are accordingly the *minority* carriers and, in heavily doped crystals, can have a ratio of a billion electrons to one. Note that some of the added electrons may occupy the vacant states (the holes) in the valence band, thus suppressing the number of holes. The generation of carriers is the basis of photovoltaic devices.

Photon Absorption

An electron can be excited into the conduction band by other means. Particularly significant is the absorption of a photon: when its energy is equal to or greater than the band gap of the specimen, the photon is absorbed and can excite an electron into the conduction band. Similar to doping, the absorption of a photon creates both majority and minority carriers. However, it is common that in photovoltaic devices, the number of light-generated carriers are much less than the number of the majority carriers already existing in the cell due to doping. Consequently, the number of majority carriers in a lighten semiconductor does not change significantly.

It is well known that semiconductors, in general, absorb radiation that falls on them to varying degrees [12]. The fraction of energy absorbed in passing through the material - the *absorption coefficient* - depends on the wavelength of the photon. A material with a low absorption coefficient only poorly absorbs incident photons, and may even appear transparent to that wavelength.

The generation of carriers due to the absorption of photons can be calculated at any location within the cell, and for any wavelength of light. However, it can be shown that the light intensity - assuming that the loss in intensity directly creates an electron-hole pair - decreases exponentially throughout the material [12]. It follows that the generation of carriers is highest at the surface. Figure (2.4) shows the generation rate of electron-hole pairs in silicon as a function of distance into the cell. There is a greater

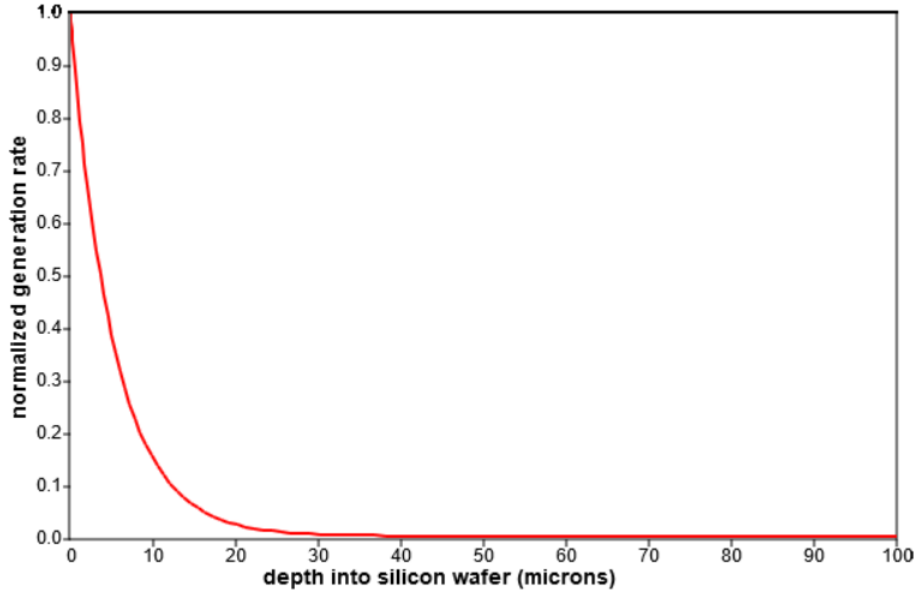


Figure 2.4: The normalized generation rate of electron-hole pairs in silicon as a function of distance into the cell [15].

generation of electron-hole pairs at the surface of the silicon cell (at 0μm), where most of the high energy light (at 700nm) is absorbed. Farther into the cell, the rate decreases exponentially and eventually becomes nearly constant.

Transport Properties

The carriers (electrons in the conduction band and holes in the valence band) in a semiconductor are free to move about the lattice that makes up the crystal structure. The carriers are, however, affected by external fields, temperature gradients, and also suffer scattering from impurities and lattice waves. The corresponding problem of transport properties, in general, is handled with the *Boltzmann equation* [12]:

$$\frac{\partial f_{\mathbf{f}}}{\partial t}\Big|_{scatt} + \frac{\partial f_{\mathbf{k}}}{\partial t}\Big|_{field} + \frac{\partial f_{\mathbf{k}}}{\partial t}\Big|_{diff} = 0. \tag{2.14}$$

It is a study of how the local concentration of carriers $f_{\mathbf{k}}(\mathbf{r})$ (in the state $\mathbf{k}$ in the neighbourhood of the point $\mathbf{r}$) changes with time t. Notice that these effects (2.14) are balanced against each other - the net rate of change of the

local carriers is zero, and there is no net overall movement in any direction.

There is much literature upon this point [16][12], but for illustrative purposes, a simple yet adequate description of carrier transport is considered here. In the semiconductor the carriers are viewed as moving in and out in random directions at some temperature dependent velocity, and occasionally collide with atoms of the crystal structure. In photovoltaic applications, the majority of the generated carriers are at the surface. This creates a concentration gradient within the semiconductor, and through random motion, the carriers are subject to the process of *diffusion*; the net movement of the carriers are from regions of high concentration to regions of low concentration. This process continues throughout the cell until the concentration becomes uniform.

Suppose that an electric field **E** is applied to the medium at a constant temperature. In its presence, the carriers move in a net direction; if the carrier is an electron, it will acquire energy and accelerate in the direction opposite to the field, while a hole will accelerate in the same direction of the field. The transport of carriers due to the presence of an electric field is called *drift transport* and does not only occur in a semiconductor material, but also in metals.

The P-N Junction

P-N junctions form not only the basis for Si solar cells but also form the basis of many other electronic devices. It is recognized as one of the essential structures in today's electronic scene which ranges from devices such as LEDs, lasers, to transistors. It is formed by joining together p-type and n-type semiconducting materials in a single crystal lattice. Since the n-type material has a high electron concentration and the p-type material a high hole concentration, joining them together causes the carriers to cross into the other regions - excess electrons diffuse from the n-type side to the p-type side, while excess holes diffuse from the p-type side to the n-type side. In this process, fixed positive ion cores (in the n-type side) and fixed negative ion cores (in the p-type side) become exposed and, as a result, establish an electric field at the junction. This forms a *depletion region* where the electric field quickly sweeps out any carriers that progress to the junction. Accordingly, the electric field tends to keep the electrons in the n-type material and the holes in the p-type material. Despite the presence of the electric field at the junction, some carriers can statistically cross the junction by diffusion.

In this event, the once labeled majority carrier becomes a minority carrier in the new region, and it will continue to diffuse away from the junction until it recombines. The current created by the diffusion of carriers across the junction is called the *diffusion current*.

A forward bias refers to the application of a voltage to the cell such that the formed electric field at the junction is reduced. By applying a positive voltage to the p-type material and a negative voltage to the n-type material, an electric field of opposite direction to that of the depletion region is established across the device. With this proviso, the carrier diffusion across the depletion region is eased and points to an increased diffusion current. This increased diffusion introduces new minority carriers at the boundary of the depletion region. The new carriers will eventually recombine with a majority carrier, and since the majority carriers are supplied from the external circuit, a net current flows under a forward bias [15].

The Photovoltaic Effect

The *light-generated current* in a solar cell involves two fundamental processes. First, an electron-hole pair must be created through the process of photon absorption, and second, the (recombination restricted) carrier must be transferred from the cell to an external circuit, where it dissipates its energy and subsequently returns to the cell.

Notice that if the carriers recombine - such that the electron-hole pair is lost - there can be no current generated. This recombination must be restricted. Fortunately, the p-n junction spatially isolates the electron-hole pair by the action of the established electric field at the junction. Under the action of a forward bias, however, a minority carrier that diffuses to the p-n junction would get swept across the boundary by the reduced electric field where it is now a majority carrier (before it can recombine).

The first generation solar cells are based on the p-n junction, starting with Si-single crystals to the adoption of bulk polycrystalline Si wafers. These cells are, to date, the most used and wide-spread technology in the market [17], and produce solar conversion efficiencies between 12% and 16% according to the manufacturing process and wafer quality [18]. They also exhibit a lifetime of about 20 years [19] which makes them suitable for roof-tops and solar farms. Despite the high conversion efficiency and long-term stability, the high costs and sophisticated manufacturing processes make it hard to be competitive with fossil fuels.

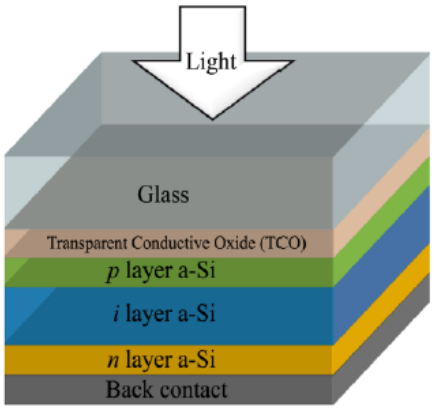


Figure 2.5: The structure of a p-i-n junction solar cell on a glass substrate.

2.2.2 Second-Generation: Thin-film technology

The second-generation solar cell is based on thin-film technology (amorphous silicon (a-Si), copper indium gallium selenide (GIGS), and cadmium telluride (CdTe)), and produce conversion efficiencies between 10% and 15%. In contrast to Si-wafer technology, the second-generation solar cell has lower material consumption in the manufacturing process, and, as such, presents a potential for a cost reduction. Unfortunately, there is still a considerable energy consumption associated with the production of the thin-film cells since it includes vacuum processes and high-temperature treatments.

A thin film is a structure with a low dimensionality that ranges from fractions of a nanometer to several micrometers in thickness. The synthesis of the material is grown step-by-step according to a well-defined deposition sequence [20] in a complicated and statistical process of particle-particle and particle-substrate interactions.

Research starting from a-Si has led to the development of thin-film solar cells in a variety of junction structures. Amorphous silicon thin-film cells, on the one hand, adopt a p-i-n structure - a three-layer sandwich-like configuration (see Figure (2.5)), while cadmium telluride adopts an n-i-p like structure. In principle, a variety of structures using different materials and depositions processes can satisfy the photovoltaic effect.

For the amorphous silicon thin-film cell, the middle layer is intrinsic silicon, while the top and bottom layers are respectively p-type (a-Si) and n-type

(a-Si). This geometry creates an electric field between the p- and n-type material and, just like the p-n junction, separates the photogenerated carriers. By using stacked structures (and different band gaps), it is possible to absorb the solar spectrum more efficiently. In any case, the focal point is, as always, to exploit the physical properties of the thin-film architecture so that the conversion efficiencies are high with a low production cost per Watt.

Modules of the second-generation solar cell have not gained the success of the first-generation because of technological problems and module stability [21]; they only amount to about 6 % of the total photovoltaic market as of 2016. However, second-generation cells can be produced on flexible substrates so that new and exciting applications can be realized.

2.2.3 Third-Generation: Organic Solar Cells

Third-generation solar cells encompass a wide range of different technologies. It may refer to the multi-junction [22], quantum dot [23], organic [24], or dye-sensitized [25] solar cells, the latter of which has been the most successful commercially. However, many promising alternatives are beginning to appear.

Dye-sensitized solar cells were developed in the early 1990s through the effort to understand and simulate the primary processes in photosynthesis [26]. They are based on a semiconductor formed between a photo-sensitized anode and an electrolyte. Dye molecules absorb light and generate positive and negative charges. It mimics the basic (3.5 billion years old well-established) process of photosynthesis: in green plants, chlorophyll molecules absorb sunlight and generate (uncollected) electric charges which are converted by redox reactions at the membrane level to generate oxygen from water and reduce carbon dioxide to carbohydrates. Conversely, in the dye-sensitized cell, stimulated electrons in the dye are rapidly injected into the conduction band, through which they diffuse to the electrode contact, and are collected as electric current.

2.3 Metal Halide Pervoskite Solar Cells

The efficient solar cell absorbs incident photons over a wide range of spectra, from visible to near-infrared wavelengths, and converts them into efficiently movable charges. The electronic structure of the efficient solar cell, then,

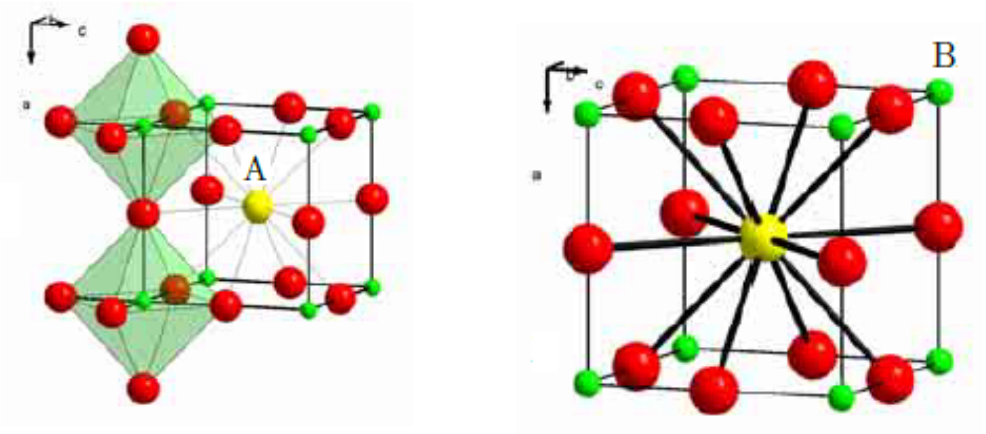
should have a suitable band gap - one that allows the absorption of different photons in the spectrum. Recently, methylammonium (MA) triiodide-plumbate and its derivatives ($CH_3NH_3PbX_3$, X=I, Br, Cl) have attracted significant interest because of their potential application as light-harvesters (and low manufacturing cost) [27].

Goldschmidt *et al* conducted much of the pioneering structural work on perovskites in the 1920s and formed the fundamental basis for further exploration of the perovskite family of compounds [28]. The corresponding structural class consists of a vast collection of compounds all having a crystal structure related to the mineral perovskite $CaTiO_3$ (ABX_3). These materials contain a large and ever-surprising variety of properties (such as superconductivity [29] and colossal magnetoresistivity [30]) and have the versatility to accommodate almost all of the elements of the periodic table. The remarkable range of structure and property interplay makes them an excellent research field for studies in physics, materials science, and solid state chemistry. In the research community of photovoltaics, however, the perovskite family of organic-inorganic halides is probably the most studied.

2.3.1 Crystal Structure

Early studies reported that the perovskite structure, ABX_3, was primarily cubic. The A cation rests at the center of the cube, occupying the 12-fold coordination site shared with twelve X anions, while the B cation sits at the cube corners in a stabilized mesh of corner sharing octahedra, as shown in Figure (2.6); the atomic coordinates are summarized in Table (2.1). As work continued, the number of proposed symmetries increased, and the research question leading the current study is formulated accordingly: What is the atomic structure of the perovskites and how do different structures influence the transport properties? In short, the exact crystal structures of $MAPbX_3$ and their underlying material properties are largely *unknown* [31].

$CH_3NH_3PbI_3$ and its derivatives are derived from ABX_3 by replacing the alkali-metal atom (A) with MA. In the highest temperature phase, $MABX_3$ crystallizes in an *average* cubic perovskite structure; the CH_3NH_3 ion is polar with C_{3v} symmetry (See Chapter 5) and is thus disordered in the cubic phase [32]. The CH_3NH_3 ion occupies 12 equivalent orientations of the C_2 axis, and the hydrogen atoms have two kinds of configurations on the C_2 axis (24 degrees of freedom) as shown in Figure (2.7a). As the temperature decreases, the framework symmetry becomes tetragonal, and the CH_3NH_3 ions



(a) (b)

Figure 2.6: The ideal cubic perovskite structure of general form ABX_3 which consists of **(a)** eight corner sharing BX_6 octahedra (only two shown) with **(b)** the A cation occupying the twelve-fold coordination site in the middle of the cube.

Site	Wyckoff Position	Fractional Coordinates
A cation	(1b)	$(\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$
B cation	(1a)	(0,0,0)
X anion	(3d)	$(\frac{1}{2},0,0)$ $(0,\frac{1}{2},0)$ $(0,0\ \frac{1}{2})$

Table 2.1: The atomic coordinates of a cubic ABX_3 perovskite structure. The lattice constant is $a = 3.905\,\text{Å}$ and the space group is $pm\bar{3}m$.

Material	**$MAPbI_3$**	**$MAPbBr_3$**	**$MAPbCl_3$**
Crystal System	Cubic	Cubic	Cubic
Transition Temperature (K)	330	236	177
Crystal System	Tetragonal	Tetragonal	Tetragonal
Transition Temperature (K)	161	149-154	172
Crystal System	Orthorhombic	Orthorhombic	Orthorhombic

Table 2.2: The crystal systems and transition temperatures for $CH_3NH_3PbX_3$, X=Cl, Br, I [31].

rotate and jump between various equivalent orientations consistent with the crystal symmetry [33], as shown in Figure (2.7b). Still further, with decreasing temperature, the tetragonal phase is transformed into the orthorhombic phase where the CH_3NH_3 ions have a definite orientation, as shown in Figure (2.7c). The crystal systems and transition temperatures are summarized in Table (2.2).

Structural Properties

Perovskites are far more complicated than traditional photovoltaic materials (such as Si) and pose significant challenges for accurate material modeling. Crystal structures are routinely studied with X-ray diffraction (XRD) techniques and also transmission electron microscopy (TEM). The atomic structure model can be constructed, in general, from the obtained crystallographic data and with the use of group theory. While the position of Bragg peaks in X-ray diffraction experiments can distinguish between the three structural phases of organic-inorganic perovskites (cubic, tetragonal, an orthorhombic), the peak intensities associated with the $(CH_3NH_3)^+$ cations are too soft to specify accurate molecular orientations. The hydrogen atoms only scatter X-radiation weakly, and so the coordinates of the H atoms cannot be resolved. Indeed, the coordinates of C and N in the $(CH_3NH_3)^+$ cation are known [33]; so too are the coordinates of PbX_3 in the orthorhombic and tetragonal phases [33][34].

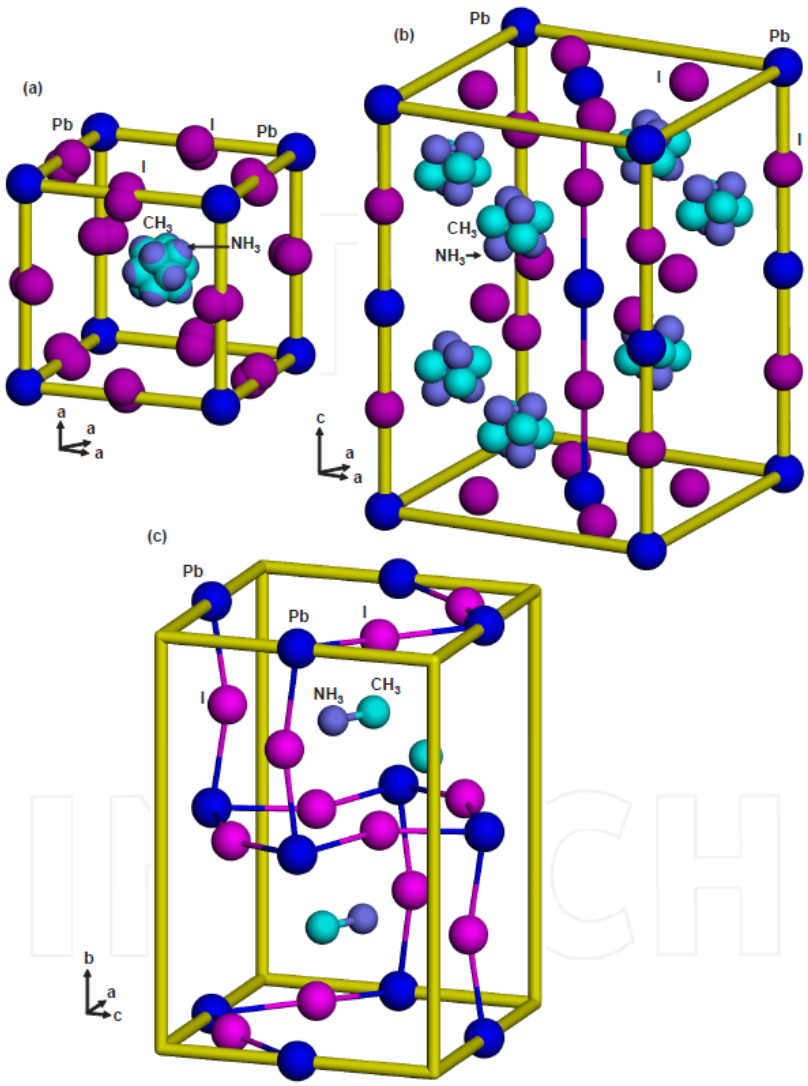


Figure 2.7: The various crystal structures of the perovskite compound [31].

Geometry Optimization

The nature of the MA cation modifies the structure of the different phases (as seen in Figure (2.7)) and influences the photovoltaic properties [35]. Therefore, to understand the *origin* of the high efficiencies exhibited by perovskite solar cells, it is essential to address the structural complexity introduced by the MA cations - it is essential to understand the crystal structures. A simplified technique, which has been applied in various studies, excludes the $(CH_3NH_3)^+$ cation altogether or replaces it by more symmetrical ones (such as Cs^+ or NH_4^+)[33][36]. Although this method is useful and has allowed researchers to investigate the band structure around the gap, it is incomplete.

The standard approach in a density functional theory calculation is to investigate the potential energy landscape associated with different molecular orientations to find a geometry that minimizes it - a configuration where the forces in the system are virtually zero [37];

$$|\Delta \mathrm{E}| \approx |\mathbf{F}||\Delta \mathrm{r}|.$$

It is for this static and balanced arrangement of atoms that the electronic structure is calculated.

The ability to efficiently *optimize* the geometry of delocalized internal coordinates and "floppy" molecules is central to perhaps the majority of all DFT calculations. Subsequently, most of the complicated structures have free internal structural parameters, and their selection considerably influences the computational effort and impacts whether the calculation will be successful or not. Although there is no need to understand in detail the operations of the optimization algorithm to make use of it, some perception of it is beneficial for many cases [38].

2.3.2 Fabrication and Performance

Perovskite solar cells originated from dye-sensitized solar cells, in which molecular dye was replaced by perovskite. Accordingly, Kojima et al. [40] demonstrated, in 2009, the use of $CH_3NH_3PbX_3$ (X = Br, I) as sensitizers in a (TiO_2) dye-sensitized solar cell structure. The perovskite was deposited onto the TiO_2 by spin coating a dimethylformamide solution containing PbX_2 and CH_3NH_3X (X = Br, I). They demonstrated power conversion efficiencies (PCE's) of around 3-4%. After two years, Im et al. [41] double the PCE by optimizing the perovskite coating conditions. However, the *ionic*

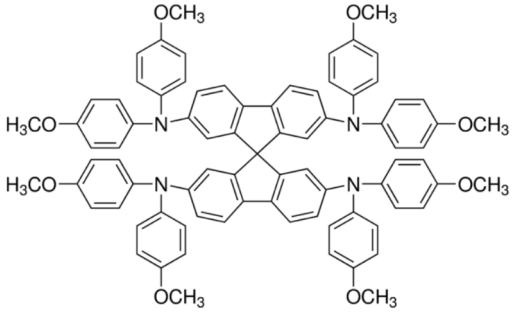


Figure 2.8: Spiro-MeOTAD is a stable and efficient hole-transport material used in solid-state dye-sensitized solar cells [39].

perovskite compounds in these devices readily dissolve in the *polar* solvent, thereby making the cells not very stable. The sensitization of TiO2 with $CH_3NH_3PbI_3$ using spiro-OMeTAD (see Figure 2.8) as the hole transporter resolved the instability problem; Kim et al. reported a long-term stable cell with a PCE of about 10% [42]. However, the corresponding scheme still produced a fairly good PCE (5.5%) in the *absence* of the hole transport material (HTM) [43]. This revealed that the perovskite can either transport holes or act as a sensitizer. In the sensitization scheme, the HTM should thoroughly infiltrate the mesoporous oxide layer (TiO_2) to induce a heterojunction. Additionally, oxide layers with electron accepting properties are thought to be needed to separate the photo-excited electrons in perovskites. To understand the charge separation in this scheme, a femtosecond transient absorption spectroscopic study was performed with an electron injecting layer TiO_2, and an insulating mesoporous layer Al_2O_3. Curiously, no considerable spectral differences were observed between TiO_2 and Al_2O_3, which means that the perovskite cell may also work in the absence of an electron injecting layer. To wit: Lee et al. confirmed the working of the perovskite cell in the absence of a TiO_2 mesoporous layer [44]. Perhaps, then, the electron transport occurs through the perovskite?

A higher PCE of 15% was realized in 2013 from a sequential (two-step) coating procedure [45]. In this approach, the $CH_3NH_3PbI_3$ layer was developed by dipping PbI_2 from solution into a mesoporous TiO_2 film into a diluted CH_3NH_3I solution. This way, the perovskite forms quickly which significantly improved the morphology of the crystal and increased the reproducibility of the cell's performance. This, in fact, pointed to a feasible

fabrication of a planar junction cell. Accordingly, since electron transfer can occur in the perovskite layer, the perovskite cell can be formed from junctions among the perovskite film (intrinsic layer), n-type thin TiO_2 film, and p-type HTM film. Indeed, the planar *p-i-n* junction device was confirmed by Liu et al. and exhibited a PCE greater than 15% [46].

2.4 Future Progress

The high-efficiency photovoltaic performance is mostly associated with the specific method of fabrication. The perovskite pigment, for instance, can be deposited in a *single-step* using a mixture of PbX_2 and CH_3NH_3X (X=Cl, Br, or I) from a common solvent. This approach generally leads to uncontrolled morphological variations which, in turn, result in weak photovoltaic performance. Therefore, a *sequential* deposition method was developed to gain better control of the composition, thickness, and morphology of the crystal. The large variety of fabrication techniques provides room for researchers to explore the fundamental physics behind the perovskite cell. However, the fundamental understanding is still largely debated and actively investigated [47][48][49].

Today, the best-performing cells have reached a PCE of 22.1% [50]. This unprecedented rise in efficiency for a photovoltaic technology - from a mere 3.8% - suggests a sunny outlook (see Figure (2.9)). However, before a large-scale deployment of the technology, some questions concerning the real potential of perovskite cells will have to be addressed.

It is well known that $CH_3NH_3PbI_3$ is unstable in humid conditions; it forms PbI_2, a carcinogen, at higher temperatures due to the loss of CH_3NH_3I. These instabilities, of course, pose real health hazards and would therefore impede outdoor applications. Lead compounds are very toxic and damaging to the environment, and so, a consortium of experts are currently investigating the health risks and dangers. With this, the solution seems simple: identify a lead-free perovskite. It should be noted that perovskite cells are still a promising and attractive photovoltaic candidate:- see Table (2.3).

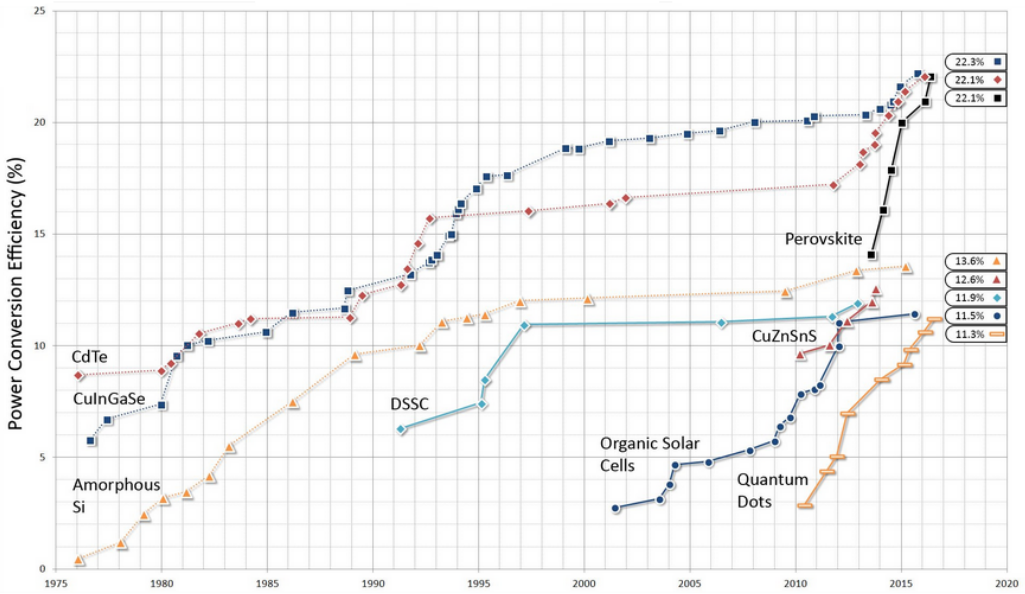


Figure 2.9: The power conversion efficiency evolution of various solar cells through the years [51].

Characteristics	CdTe	CIGS	Si	Perovskite
Raw Material Cost	Low	Medium	Low	**Low**
Finished Material Cost	Low	High	High	**Low**
Fabrication Cost	Medium	Medium	High	**Low**
Energy Payback Period	Medium	High	High	**Low**
Efficiency	Medium	Medium	High	**High**

Table 2.3: The current advantages of the perovskite solar cell [26].

Chapter 3

Density Functional Theory

3.1 What is Density Functional Theory?

One of the most thorough advances of the twentieth century was the development of quantum mechanics. Almost all subatomic elements of matter and their interactions are welcomed within its framework, and in principle, can be described exactly. This is the subject of condensed matter physics and materials science: understanding the properties of interacting electrons and nuclei. Unfortunately, the countless collection of electrons and nuclei that make up the material under study are strongly interacting, and as a result, makes the direct solution of the Schrödinger equation extremely difficult. In point of fact, the alleged many-body system is practically impossible to solve exactly. It is therefore understood that progress depends on the development of approximate and tractable solutions to the many-body problem.

The technological base is becoming progressively dependent on complicated "real" materials, and so it is natural that there is a keen interest in tractable methods for solving the many-body problem. Density functional theory (DFT) is a remarkably successful quantum mechanical approach to this problem. It is a computational modelling method used to investigate the electronic structure of many body systems. In chemistry, for instance, density functional theory is routinely applied for calculating the binding energy of molecules, while in physics, the band structure of solids. It is an extremely versatile method and is consequently quickly becoming a standard tool for materials modeling problems.

3.2 The Schrödinger Equation

In quantum mechanics, it is understood that all perceptible information of a given system is contained in the system's wavefunction ψ. Therefore, understanding and extracting quantum mechanical information amounts to solving the Schrödinger equation. Given the fact that this work is exclusively concerned with the electronic structure of atoms, molecules, and solids, a solution to the non-relativistic time-independent Schrödinger equation is desired. For a single electron moving in a potential $V(\mathbf{r})$, the corresponding Schrödinger equation translates to

$$\left[-\frac{\hbar^2}{2m}\nabla^2 + V(\mathbf{r})\right]\psi = \epsilon\psi. \tag{3.1}$$

The bracket terms describe, respectively, the kinetic and potential energies of the electron. Notice that the detailed definition of the bracket terms (the Hamiltonian) depend on the physical system. Companions to the Hamiltonian are the wavefunctions ψ_n, a set of solutions called eigenstates. Each of these solutions has an associated eigenenergy ϵ_n that satisfy the eigenvalue problem.

There are certainly well-known examples for which the Schrödinger equation can be solved exactly. A trivial form of the equation is that for a single electron confined to a length $x = L$ by infinite barriers. The potential is given by

$$V(x) = \begin{cases} \infty & x \leq 0, x \geq 0 \\ 0 & 0 < x < L \end{cases} \tag{3.2}$$

which yields for the wave equation

$$-\frac{\hbar^2}{2m_e}\frac{d^2\psi(x)}{dx^2} = \epsilon\psi(x). \tag{3.3}$$

In this case, m_e, $\psi(x)$, and ϵ are respectively the mass, wavefunction, and energy of the electron. By reason of physical properties, the electron's kinetic energy must be finite and so the particle can never penetrate the infinite potential barriers. Consequently, there is a zero probability for the particle to be located at $x \leq 0$ or $x \geq L$. In order to account for this result, the wavefunctions must satisfy the boundary conditions $\psi(0) = 0$ and $\psi(L) = 0$. Consequently, it can be shown that the wavefunctions must be sinelike

$$\psi_n(x) = A \sin\left(\frac{n\pi}{L}x\right) \qquad n = 1, 2, 3, ... \tag{3.4}$$

The possible energies are found by substituting the wavefunction into the Schrödinger equation and solving for ϵ. The result is

$$\epsilon_n = \frac{\hbar^2}{2m}\left(\frac{n\pi}{L}\right)^2, \tag{3.5}$$

where n is the principle quantum number. Notice that the quantization of the energy occurs in a natural way from the application of the boundary conditions to the wavefunction.

It is instinctive to extend this example to three dimensional space seeing that it is where all known matter exists. Therefore, the above parallel example is that of an electron trapped in a box with planar faces at $x = 0$, $x = L$, $y = 0$, $y = L$, $z = 0$, and $z = L$. If the electron is free, then the wave equation is similarly given by

$$-\frac{\hbar^2}{2m_e}\nabla^2\psi(x, y, z) = \epsilon\psi(x, y, z) \tag{3.6}$$

where the three dimensional wavefunction is subject to the boundary condition $\psi = 0$ on the walls of the box. Notice that this boundary condition is the three dimensional generalization of equation (3.2).

In this illustration, the mathematical solution will follow from the physics. Since the electron is free within the box, the $x-$, $y-$, and $z-$ dependent part of the wavefunction must be independent of each other. Therefore, by the assumption that the wavefunction can be represented by the product of three independent functions (one for each dimension), the solution has the following form

$$\psi(x, y, z) = X(x)Y(y)Z(z). \tag{3.7}$$

Substituting the assumed solution into equation (3.6) and division of the result by (3.7) gives

$$-\left(\frac{1}{X(x)}\frac{d^2X}{dx^2} + \frac{1}{Y(y)}\frac{d^2Y}{dy^2} + \frac{1}{Z(z)}\frac{d^2Z}{dz^2}\right) = \frac{2m_e}{\hbar^2}\epsilon \tag{3.8}$$

where each term involves a function of only one variable. If equation (3.8) is to hold for any value of the independent variables, each of the three

terms must be separately constant. In that respect, the separation in the x-coordinate yields

$$\frac{d^2X}{dx^2} = -k_x^2 X(x) \tag{3.9}$$

with trivial solution $X(x) = A\sin(k_x x) + B\cos(k_x x)$. Applying the boundary conditions at $x = 0$ and $x = L$, the solution is reduced to

$$X(x) = \sin\left(\frac{n_x \pi}{L}\right) \qquad n_x = 1, 2, 3, ... \tag{3.10}$$

with $k_x L = n_x \pi$. Similar treatment of the Y and Z equations with separation constants $-k_y$ and $-k_z$ yields

$$Y(y) = \sin\left(\frac{n_y \pi}{L}\right) \qquad n_y = 1, 2, 3, ... \tag{3.11}$$

$$Z(z) = \sin\left(\frac{n_z \pi}{L}\right) \qquad n_z = 1, 2, 3, ... \tag{3.12}$$

with $k_y L = n_y \pi$ and $k_z L = n_z \pi$. Therefore, according to equation (3.7), the total wavefunction of a free particle in a box is written as a product of the separable solutions. That is,

$$\psi(x, y, z) = A\sin\left(\frac{n_x \pi}{L}\right)\sin\left(\frac{n_y \pi}{L}\right)\sin\left(\frac{n_z \pi}{L}\right) \tag{3.13}$$

with n_x, n_y, and n_z being positive integers. On that account, the boundary value problem will not have a solution for arbitrarily chosen values of ϵ. Instead, it can be shown that the ϵ values must satisfy

$$\epsilon_n = \frac{\pi^2 \hbar^2}{2m_e}\left(\frac{n_x^2}{L^2} + \frac{n_y^2}{L_2^2} + \frac{n_z^2}{L^2}\right) \tag{3.14}$$

with positive integer values for n_x, n_y and n_z.

In these examples, the Schrödinger equation is solved exactly, and as a result, valuable intuition about the quantum mechanical system is gained: The electron can only occupy certain positive energy levels and is more likely to be found at certain positions than others. The exact intelligible results of the "particle in a box" problem is a consequence of the simplistic nature of the potential $V(\mathbf{r})$. It is zero in these examples and the Schrödinger equation reduces to (3.6). In practice, however, it is relevant to study systems in which multiple electrons interact with multiple nuclei. The potential is thus

recognized as a complicated functional form of positions. This means that the Schrödinger equation is generally very difficult to solve, and in point of fact, usually amounts to solving elaborate approximations of the Schrödinger equation. Despite the intrinsic uncertainty that results from approximate solutions, they are often good enough to make powerful predictions about the properties of a complicated, many-body system.

3.2.1 The Many-Body Problem

An atom consists of a dense nucleus of protons and neutrons about which a probabilistic distribution of electrons is centered. When an enormous collection of atoms are brought together to form a solid, the electron orbitals overlap and interact in such a way that generally establishes a minimum energy configuration. If we want to outline characteristic properties of these well-defined collections of electromagnetically interacting atoms, the fundamental piece of information is, indubitably, the wavefunction of the system. However, when many particles are present, the prevailing overlap is so considerable that the quantization scheme in which each electron is localized on its own atom must breakdown [16]. As a consequence, the wavefunction is a function of the coordinates of all the particles

$$\psi = \psi(\mathbf{r}_1, ..., \mathbf{r}_n; \mathbf{R}_1, ..., \mathbf{R}_N). \tag{3.15}$$

This means that the wavefunction of a given particle cannot be found without simultaneously considering the individual wavefunctions associated with all the other particles. Therefore, if a given crystal structure has N nuclei with $N + ZN$ interacting particles (where Z is the atomic number), the corresponding full wavefunction is a $3(N + ZN)$-dimensional function. In view of the fact that the "many-body" wavefunction cannot be decoupled - as is the case in the three dimensional free particle - it is virtually impossible to find.

The Hamiltonian for the whole system of ions and electrons of the assumed crystal is given by

$$
\begin{aligned}
H = & -\frac{\hbar^2}{2}\sum_i \frac{\nabla^2_{\mathbf{R}_i}}{M_i} - \frac{\hbar^2}{2}\sum_i \frac{\nabla^2_{\mathbf{r}_i}}{m_e} \\
& - \frac{1}{4\pi\epsilon_0}\sum_{i,j}\frac{e^2 Z_i}{|\mathbf{R}_i - \mathbf{r}_j|} + \frac{1}{8\pi\epsilon_0}\sum_{i\neq j}\frac{e^2}{|\mathbf{r}_i - \mathbf{r}_j|} + \frac{1}{8\pi\epsilon_0}\sum_{i\neq j}\frac{e^2 Z_i Z_j}{|\mathbf{R}_i - \mathbf{R}_j|}
\end{aligned}
\tag{3.16}
$$

where M_i is the mass of the nucleus at $\mathbf{R}_i$, and m_e that of the electron located at $\mathbf{r}_i$. The first two terms are the kinetic energy operators for the nuclei and electrons respectively. The three remaining terms, in order, describe the Coulomb interaction between electrons and nuclei, between electrons and electrons, and between nuclei and nuclei. In agreement with the above discussion, it is impossible to solve this problem exactly because of the enormously complicating effects of the interactions between the particles. Therefore, further advancements demand some simplifying physical ideas.

Born-Oppenheimer Approximation

A basic observation is that atomic nuclei are much heavier than individual electrons. The ratio of the mass for each proton or neutron in a nucleus to an electron is approximately 2000. This means that a single proton, for instance, is three orders of magnitude heavier than the electron. Because of this difference, the electrons move much faster than the nuclei. Also, in consideration of their opposite charge, there are mutual attractive forces acting between the atomic nuclei and the electrons. These forces have the effect of accelerating the particles, and since the acceleration is inversely proportional to the mass, $a = f/m$, the acceleration of the electrons is considerably larger than that of the nuclei. This means that electrons are moving and answering to forces much more promptly in comparison to the nuclei. Therefore, it is credible to assume that on the typical time-scale of nuclear motion, electrons will respond almost instantaneously and relax to their ground-state configuration. In this sense, the nuclei are approximated to be fully localized about a single point. Therefore, we say that the physical system is split into two pieces: the electronic part and the nuclear part,

$$
\Psi = \psi_e(\mathbf{r}_1, ..., \mathbf{r}_n; \mathbf{R}_1, ..., \mathbf{R}_N)\psi_n(\mathbf{R}_1, ..., \mathbf{R}_N). \tag{3.17}
$$

This separation is the great Born-Oppenheimer approximation [52]. The electronic part ψ_e satisfies the Schrödinger equation for electrons in a static

lattice, while the nuclear part ψ_n satisfies a Schrödinger equation of the ions alone. Notice that all physical properties of the electrons depend parametrically on the given nuclear positions.

In view of the the Born-Oppenheimer approximation, the kinetic energy term corresponding to the nuclei in the Hamiltonian (3.16) vanishes while the Coulomb interaction between nuclei and nuclei (the last term) becomes a constant. Therefore, the system of $N + ZN$ interacting particles carefully reduces to that of ZN interacting electrons moving in the potential of the nuclei. The Hamiltonian is given by

$$H = -\frac{\hbar^2}{2}\sum_i \frac{\nabla^2_{\mathbf{r}_i}}{m_e} - \frac{1}{4\pi\epsilon_0}\sum_{ij}\frac{e^2 Z_i}{|\mathbf{R}_i - \mathbf{r}_j|} + \frac{1}{8\pi\epsilon_0}\sum_{i\neq j}\frac{e^2}{|\mathbf{r}_i - \mathbf{r}_j|}. \tag{3.18}$$

From a fundamental point of view, the resulting Schrödinger equation with (3.18) as the Hamiltonian is still hopelessly insoluble. The pending problem of solvability is a result of the fact that the Born-Oppenheimer approximation does not develop the Coulomb interaction between electrons and electrons in such a way that the full wavefunction reduces to a more tractable form. By reason of preceding explanation, the wavefunction, therefore, persists as a complicated function of the distance between the electrons (3.15).

In fairness to the Born-Oppenheimer approximation, it merely debates the dynamical problem between conduction electrons and nuclei; not the electron-electron interactions. Therefore, for purposes of studying electron dynamics, the underlying problem of solving the many-body Schrödinger equation is now a question of how to represent the effects of the electron-electron interactions.

3.2.2 Wavefunction Based Methods

In keeping with the above discussion, the Schrödinger equation reflecting the Born-Oppenheimer approximation is given by

$$-\frac{\hbar^2}{2}\sum_i^N \frac{\nabla^2_{\mathbf{r}_i}}{m_e}\Psi - \frac{1}{4\pi\epsilon_0}\sum_{ij}^N \frac{Z_i e^2}{|\mathbf{R}_i - \mathbf{r}_j|}\Psi + \frac{1}{8\pi\epsilon_0}\sum_{i\neq j}\frac{e^2}{|\mathbf{r}_i - \mathbf{r}_j|}\Psi = E\Psi. \tag{3.19}$$

The second term represents the attractive electrostatic potentials between the fixed nuclei and electrons at sequential points $\mathbf{R}_i$ and $\mathbf{r}_j$, and the last term represents the challenging electrostatic interactions of the individual electrons.

The Hartree Equations

Consistent with the enduring theme of making relevant approximations to demonstrate the power of the Schrödinger equation, the electron-electron interactions are crudely interpreted as that of a single electron interacting with the average field of the others. Development of this interpretation leads to the conclusion that a single electron does not perceive the instantaneous effects of the other electrons, and as a result, has entirely independent and *uncorrelated* motion [12]. This allows treatment of the complete wave function as separable factors, each corresponding to the coordinates of a single electron

$$\Psi(\mathbf{r}_1 s_1, \mathbf{r}_2 s_2, ..., \mathbf{r}_N s_N) = \psi_1(\mathbf{r}_1 s_1)\psi_2(\mathbf{r}_2 s_2)...\psi_N(\mathbf{r}_N s_N). \tag{3.20}$$

Give notice to the fact that this treatment is understood as an exploitation of the solvability of the one-particle Schrödinger equation. At first glance, the assumption that any given electron neglects its explicit interaction with the field of the others is very unconvincing - especially in consideration of the fact that electrons are related to each other by the Coulomb repulsion - it admittedly leads to a simple theory from which quantitative results can be deduced. In this respect, the argument of uncorrelated electron motion is acknowledged, but it is assumed to be not entirely correct. We will explore this in more detail below.

The corresponding set of the one-electron Schrödinger equation - called the Hartree equations - is given by

$$-\frac{\hbar^2}{2m_e}\nabla^2\psi_i(\mathbf{r}) - \frac{1}{4\pi\epsilon_0}\sum_{\mathbf{R}}\frac{Ze^2}{|\mathbf{R}-\mathbf{r}|}\psi_i(\mathbf{r}) + U_e(\mathbf{r})\psi_i(\mathbf{r}) = \epsilon_i\psi_i(\mathbf{r}), \tag{3.21}$$

where $U_e(\mathbf{r})$ is the approximated electron-electron interactions. As above-mentioned, $U_e(\mathbf{r})$ describes the Coulomb repulsion between a single electron and the total electron density

$$n(\mathbf{r}) = -e \sum_i |\psi_i(\mathbf{r})|^2 \tag{3.22}$$

and is given by,

$$U_e(\mathbf{r}) = -e \int \frac{n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' = e^2 \sum_i \int \frac{|\psi_i(\mathbf{r})|^2}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}'. \tag{3.23}$$

This equation is formally known as the Hartree potential [12], and is one of the many attempts of treating electron-electron interactions. Notice that the perception of uncorrelated electron motion follows from the fact that the electron density $n(\mathbf{r})$ itself is interpreted as an arbitrary particle that interacts with a single electron. Accordingly, the system reduces to a two-body problem that is consistent with the idea of uncorrelated motion and equation (3.20).

As crude as this approximation is to the full Schrödinger equation, the mathematical task of finding ψ_i is still of considerable complexity. For instance, to solve the Hartree equations (3.21), the Hartree potential $U_e(\mathbf{r})$ must be defined. To define the Hartree potential the electron density must be known, except to know the electron density, the Hartree equations must be solved. Therefore, ψ_i must currently be known (despite the fact that it is not), to find it. In practice, this paradox is broken by iteration: choose a potential so that ψ_i can be found, compute a new potential from the resulting wavefunction, and find a better ψ_i. This procedure continues until the resulting potential is *self-consistent*, meaning that further iterations do not change the potential.

A fundamental mistake in the Hartree approximation is exposed if we refer to the electron wavefunction (3.20) and note that it is a function of not only coordinates, but also spin variables. This means that we need to take into account the Pauli exclusion principle which requires that the true electron wavefunction be antisymmetric with respect to switching or *exchanging* any of two electrons [12]:

$$\Psi(\mathbf{r}_1 s_1, \mathbf{r}_2 s_2, ..., \mathbf{r}_N s_N) = -\Psi(\mathbf{r}_2 s_2, \mathbf{r}_1 s_1, ..., \mathbf{r}_N s_N). \tag{3.24}$$

Notice that in excluding the uninteresting case in which Ψ vanishes identically in equation (3.20), the antisymmetry requirement (3.24) cannot be satisfied for the product configuration for the electron wavefunction (3.20). To that end, the true electron wavefunction thus incorporates somewhat of

a correlation effect - which is conceptually different than that caused by the Coulomb repulsion - that influences only electrons of like spin. Accordingly, it is said that the Hartree approximation suffers from the lack of electronic *exchange*.

Hartree-Fock Theory

The lack of exchange in the Hartree approximation is accounted for by replacing the simple product wavefunction (3.20) by a Slater determinant [53]:

$$\Psi(\mathbf{r}_1 s_1, \mathbf{r}_2 s_2, ..., \mathbf{r}_N s_N) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \psi_1(\mathbf{r}_1 s_1) & \psi_1(\mathbf{r}_2 s_2) & \dots & \psi_1(\mathbf{r}_N s_N) \\ \psi_2(\mathbf{r}_1 s_1) & \psi_2(\mathbf{r}_2, s_2) & \dots & \psi_2(\mathbf{r}_N s_N) \\ \vdots & \vdots & & \vdots \\ \psi_N(\mathbf{r}_1 s_1) & \psi_N(\mathbf{r}_2 s_2) & \dots & \psi_N(\mathbf{r}_N s_S) \end{vmatrix},$$

where the single particle wavefunctions are subject to the orthonormal constraint

$$\int \psi_i^*(\mathbf{r})\psi_j(\mathbf{r})d\mathbf{r} = \langle \psi_i | \psi_j \rangle = \delta_{ij}. \tag{3.25}$$

Notice that this ansatz trivially satisfies the antisymmetric property of the exact solution since exchanging variables simply exchanges two columns, and therefore, reverses the sign of the whole. Additionally, it is clearly true that Ψ vanishes identically when any two of the functions are identical which is a clear depiction of the Pauli principle.

It can be shown that the expectation value of the Hamiltonian (3.21) in the suitable ansatz state given by the Slater determinant is

$$\begin{aligned} \langle H \rangle = &\sum_i \int d\mathbf{r} \psi_i^*(\mathbf{r}) \left(-\frac{\hbar^2}{2m_e}\nabla^2 + U_{ion}(\mathbf{r}) \right) \psi_i(\mathbf{r}) \\ &+ \frac{1}{2}\sum_{i,j} \int d\mathbf{r} d\mathbf{r}' \frac{e^2}{|\mathbf{r}-\mathbf{r}'|} |\psi_i(\mathbf{r})|^2 |\psi_j(\mathbf{r}')|^2 \\ &- \frac{1}{2}\sum_{i,j} \int d\mathbf{r} d\mathbf{r}' \frac{e^2}{|\mathbf{r}-\mathbf{r}'|} \delta_{s_i s_j} \psi_i^*(\mathbf{r})\psi_i(\mathbf{r}')\psi_j^*(\mathbf{r}')\psi_j(\mathbf{r}), \end{aligned} \tag{3.26}$$

where $U_{ion}(\mathbf{r})$ is the potential of the ions and is given by

$$U_{ion}(\mathbf{r}) = -\frac{1}{4\pi\epsilon_0}\sum_{\mathbf{R}}\frac{Ze^2}{|\mathbf{R}-\mathbf{r}|}. \tag{3.27}$$

Accordingly, equation (3.26) is known as the Hartree-Fock energy. Notice that the last term (called the exchange energy) is of significant interest on account of the fact that it embodies the antisymmetric requirement imposed by the Pauli principle - it vanishes when $s_i \neq s_j$.

The ground state wavefunctions $\psi_i(\mathbf{r})$ are determined by minimizing the Hartree-Fock energy expression under the constraint that the single electron wavefunctions are orthonormal (3.25). In this respect, the minimisation problem is realized by using the Euler-Lagrange method. The resulting stationary condition is given by:

$$\delta\left(\langle H_0\rangle - \sum_{ij}\epsilon_{ij}(\langle\psi_i|\psi_j\rangle - 1)\right) = 0, \tag{3.28}$$

where $\langle H_0\rangle$ is the Hartree-Fock ground state energy and ϵ_{ij} designates the Lagrange multipliers. Contingent upon this procedure, the Hartree-Fock equations are given by:

$$\begin{aligned} -\frac{\hbar}{2m_e}\nabla^2\psi_i(\mathbf{r}) + U_{ion}(\mathbf{r})\psi_i(\mathbf{r}) + U_e(\mathbf{r})\psi_i(\mathbf{r}) \\ -\sum_j\int d\mathbf{r}'\frac{e^2}{|\mathbf{r}-\mathbf{r}'|}\psi_j^*(\mathbf{r}')\psi_i(\mathbf{r}')\psi_j(\mathbf{r})\delta_{s_i,s_j} = \epsilon_i\psi_i(\mathbf{r}) \end{aligned} \tag{3.29}$$

which is a generalization of the Hartree equations.

Notice that the Hartree-Fock equations (3.29) is distinguished from the Hartree equations (3.21) by one term on the left side - the exchange term which has the structure of an integral operator:

$$\int V_x(\mathbf{r},\mathbf{r}')\psi_i(\mathbf{r}')d\mathbf{r}' = -\sum_j\int d\mathbf{r}'\frac{e^2}{|\mathbf{r}-\mathbf{r}'|}\psi_j^*(\mathbf{r}')\psi_i(\mathbf{r}')\psi_j(\mathbf{r})\delta_{s_i,s_j}. \tag{3.30}$$

The intricacy involved in the exchange term induces additional complexity to the task of finding ψ_i. For this reason, the Hartree-Fock equations cannot, in general, be solved analytically. One exception is the free electron gas in

which the assumption of a periodic potential (zero or constant) leads to a set of exact plane wave solutions. In other situations, the Hartree-Fock equations are solved self-consistently as was the case for the Hartree equations.

3.2.3 Discussion

In view of the fact that the problem of electrons in a solid is a many-body problem, serious approximations must be formed in exchange for information about the system. An intuitive foundation upon which approximations are built is to completely localize the massive nuclei about fixed points, and allow "free communications" between the electrons and the fixed nuclei. This is the so-called Born-Oppenheimer approximation, and in consideration of pair potentials describing electron-electron interactions, it insufficiently simplifies the many-body problem to a tractable one. Therefore, a lot of effort has been expended on the problem of how best to approximate electron-electron interactions.

In this respect, Hartree proposed that the classical electron-electron interactions can somewhat be represented by a clever choice of a mean-field potential. As a result, every electron was regarded to move in an effective single-particle potential,

$$V_{eff}(\mathbf{r}) = -V_{ext}(\mathbf{r}) + V_H(\mathbf{r}), \tag{3.31}$$

where V_{ext} and V_H have assumed the former role of U_{ion} and U_e respectively. This leads to a set of self-consistent single-particle equations,

$$\left[-\frac{\hbar^2}{2m}\nabla^2 + V_{ext}(\mathbf{r}) + V_H(\mathbf{r})\right]\psi_i(\mathbf{r}) = \epsilon_i\psi(\mathbf{r}), \tag{3.32}$$

which approximately describe the electronic structure of atoms. The problem with this approximation is that it does not reflect electronic exchange interactions. This matter of contention - the lack of exchange - is settled by Hartree-Fock theory. Accordingly, the product wavefunction in the Hartree formalism is replaced by a Slater determinant.

Notice that Hartree, and by extension, Hartree-Fock theory, are not exact theories. In real systems, exact wavefunctions cannot, in general, be expressed by single Slater determinants. The single-determinant approximation does not take into account the Coulomb correlated motion of the electrons. As a result, the probability of finding two electrons close together

is much less than what should be expected on the basis of the uncorrelated motion of Hartree-Fock [16]. This has the general effect of reducing the electrostatic energy, and in this sense, articulates an inherent mistake in the mean-field approximation.

The interaction energy missed by Hartree-Fock is recognized as the correlation energy E_c,

$$E_c = E_0 - E_{HF}, \tag{3.33}$$

where E_0 is the true ground state energy and E_{HF} the Hartree-Fock ground state energy. Given that the Hartree-Fock method is variational,

$$E_{HF} \geq E_0,$$

the correlation energy is strictly a negative quantity. The only exception is that for a single electron system in which case $E_c = 0$.

At this point, it is clear that improvements to the many-body problem must incorporate correlation effects outside the limits of Hartree-Fock. Such methods (configuration interaction [54], coupled-cluster [55] and Møller-Plesset theory [56]) are called post Hartree-Fock, and have been carefully developed in quantum chemistry to include correlation effects. It should be noted, however, that the best correlated methods are restricted to small interacting systems.

3.3 Density Functional Theory

Density Functional Theory (DFT) is a ground state theory that provides clear insight into the many-body problem of Section (3.2.1). DFT is different than wavefunction based methods because it considers as the elementary variable the electronic charge density. The significant interest placed on the charge density is largely motivated by the accentuated concern of the unattainable wavefunction solution for any but the simplest systems. To that end, it will be shown that when working with the charge density - instead of the wavefunction - one can efficiently extract information from relatively large systems.

3.3.1 The Hohenberg-Kohn Theorems

At the core of DFT are two powerful mathematical theorems developed in 1964 by Hohenberg and Kohn [57]. They formally express the electron density as the central variable describing electron interactions. The first theorem can be stated as follows: The total ground state energy E of the Schrödinger equation is a unique functional of the electron density $n(\mathbf{r})$.

The notably simple proof of the first theorem, as depicted by Hohenberg and Kohn, proceeds by *reductio ad absurdum*. Recall that we are strictly concerned with a collection of an arbitrary number of electrons subject to their mutual Coulomb repulsion and the potential due to the fixed nuclei. In accordance, the Hamiltonian is given by equation (3.18), which can be compactly written as:

$$\hat{H} = \hat{F} + \hat{V}_{ext} \tag{3.34}$$

where the first term is the universal electronic Hamiltonian composed of the kinetic energy operator $\hat{T}$ and the interaction operator $\hat{V}$,

$$\hat{F} = \hat{T} + \hat{V}. \tag{3.35}$$

The last term in equation (3.34) is the electron-nuclei interaction operator, where the subscript "ext" is used to refer to the fact that the positive charges of the nuclei are *external* to the system of electrons.

To persist in the proof, we assume by design that there are two different external potentials, $V_{ext,1}(\mathbf{r})$, and $V_{ext,2}(\mathbf{r})$ that give rise to the same electron density $n(\mathbf{r})$. Associated with the distinct external potentials are two separate Hamiltonians, $\hat{H}_1$ and $\hat{H}_2$, that respectively correspond to different ground state wavefunctions, Ψ_1 and Ψ_2. If we assume that the ground state is not degenerate, then in view of the minimal property of the ground state energy, there is a comparative restriction between E_1 and E_2. Accordingly, the restriction is exactly expressed as:

$$E_1 = \langle \Psi_1 | \hat{H}_1 | \Psi_1 \rangle < \langle \Psi_2 | \hat{H}_1 | \Psi_2 \rangle . \tag{3.36}$$

Given that the present ground state densities, $n(\mathbf{r})$, are assumed to be identical, the right hand side of equation (3.36) can be written as:

$$\begin{aligned}\langle\Psi_2|\,\hat{H}_1\,|\Psi_2\rangle &= \langle\Psi_2|\,\hat{H}_2\,|\Psi_2\rangle + \langle\Psi_2|\,\hat{H}_1 - \hat{H}_2\,|\Psi_2\rangle \\ &= E_2 + \int n(\mathbf{r})\,[V_{ext,1}(\mathbf{r}) - V_{ext,2}(\mathbf{r})]\,d\mathbf{r}.\end{aligned} \tag{3.37}$$

where the bracket terms follow from the universality of $\hat{T}$ and $\hat{V}$. Notice that an expression for equation (3.37) that is equally well-founded is that for which the subscripts are reversed:

$$\langle\Psi_1|\,\hat{H}_2\,|\Psi_1\rangle = E_1 + \int n(\mathbf{r})\,[V_{ext,2}(\mathbf{r}) - V_{ext,1}(\mathbf{r})]\,d\mathbf{r}. \tag{3.38}$$

Adding equations (3.36) and (3.37) gives

$$E_1 + E_2 < E_2 + E_1 \tag{3.39}$$

which is an unequivocal contradiction. As a result, this indicates that the assumption of the existence of two external potentials that give rise to the *same* ground state electron density $n(\mathbf{r})$ is not correct. It follows that the ground state density uniquely determines the external potential (to within an additive constant), and by extension, all the the properties of the system derivable from the Schrödinger equation. With this proviso, the total ground state energy E is indeed a unique functional of the electron density $n(\mathbf{r})$ and can conveniently be written as:

$$E[n(\mathbf{r})] = F[n(\mathbf{r})] + \int n(\mathbf{r})V_{ext}(\mathbf{r})d\mathbf{r}, \tag{3.40}$$

where $F[n(\mathbf{r})]$ is the universal - though unknown - expectation value of the electron density defined in equation (3.35).

The second Hohenberg-Kohn theorem states that the density that minimises the total energy (3.40) is the exact ground state density. It is commonplace to regard the second theorem as obtaining the ground state energy variationally. The proof is straightforward. If we comprehensively enlarge the scope of the first theorem to include the fact that there is an implicit correspondence between the electronic ground state density and the corresponding wavefunction, then we can write

$$E_0 = E_0[n(\mathbf{r})] = \langle\Psi|\,\hat{H}\,|\Psi\rangle\,. \tag{3.41}$$

Incidentally, if we do not know $n(\mathbf{r})$ exactly, then the variational principle demands that any trial state Ψ' corresponding to $n'(\mathbf{r})$ give rise to a necessarily higher or equal energy:

$$E[n'(\mathbf{r})] = \langle \Psi' | \hat{H} | \Psi' \rangle \geq E_0[n(\mathbf{r})]. \tag{3.42}$$

For non-degenerate ground-states, equality holds only if Ψ' is the ground state wavefunction corresponding to the exact ground state density $n(\mathbf{r})$. Therefore, the ground state density that is in accordance with the minimum energy is the exact ground state density.

A transparent matter of uncertainty remains - why do we explicitly make use of the electron density [58]? Theoretical work that follows the path of the Schrödinger equation is accustomed to express the theory of electronic structure in terms of single-particle wavefunctions. For instance, the Hartree-Fock formalism of Section (3.2.1) treats the complete wavefunction as separable factors, each corresponding to a single electron. As a result, if we study a large system of interacting particles, then many Slater determinants (3.2.2) are required which effectively obscures qualitative (and quantitative) comprehension. To that end, DFT provides a complementary perspective. It concentrates principally on the electron density which is actually physically observable, and on this account, can be easily visualized - even for large systems. It therefore provides enhancing insight into the nature of the many-body problem in a way that wavefunction methods never can.

3.3.2 The Challenge of the Hohenberg-Kohn Theorems

The results of the Hohenberg-Kohn theorems allows for the problem of solving the many-body Schrödinger equation for non-degenerate ground-states to be cast into a variational problem of minimising the energy functional $E[n(\mathbf{r})]$ with respect to the electron density in a given external potential. Whether or not this problem is easier than the formidable task of solving for the wavefunction from the full Schrödinger equation is yet to be determined. For the moment, however, we can assume that DFT is a compliant method in comparison to wavefunction based methods. This is sensible on account of the fact that instead of dealing with a function of $3N$ variables (the many-body wavefunction), we are now dealing with a function of only three variables (the electron density). The complexity of the problem has, therefore, rationally been reduced.

In the interest of gathering physical information that can be used to make contributions to a diverse range of scientific questions, DFT appoints significant emphasis to the fact that there is a one-to-one correspondence between the electron density $n(\mathbf{r})$ and the external potential $V_{ext}(\mathbf{r})$. Since the external potential depends on a set of parameters - the positions of the nuclei - the energy functional (3.40) can, in principle, be minimized with respect to *these* parameters. This allows for direct physical predictions of a crystal structure [37]. Specifically, DFT can predict lattice constants, molecular geometries, charge distributions, and so on, of a given crystal structure. The overarching point of the inherent predictable power of DFT lends itself to diverse material modelling problems in general. Accordingly, it can predict phase diagrams, transport properties, thermodynamic properties, mechanical properties (elastic modulus, compressibility, thermal expansion coefficients), magnetic properties, and chemical properties of a given material.

Notice that we are briefly avoiding the prominent details of how DFT calculations can actually yield physical information. This is because the enclosing point of the section is to indicate that the Hohehberg-Kohn theorems, in fact, do not actually offer a practical way to calculate the ground state density of a system, and by extension, any of the aforementioned physical properties of a given material. The Hohenberg-Kohn theorems do not formally express in any way what the energy functional, $E[n(\mathbf{r})]$, actually is. They purely prove that a functional of the electron density exists that can, in principle, be used to solve the Schrödinger equation. Therefore, one year after the seminal DFT publication, Kohn and Sham [59] constructed a simple method to perform DFT calculations that preserve the exact nature of DFT. This is the subject of interest for next section.

3.3.3 The Kohn-Sham Formulation

The underlying and distinctive purpose of the Kohn-Sham formulation is to map the interacting system of electrons onto a fictitious non-interacting system of electrons. This, as a conclusion, allows us to find the true minimizing density $n(\mathbf{r})$ by solving a set of single-particle Schrödinger equations [59]:

$$\left[-\frac{\hbar^2}{2m}\nabla^2 + V_{ext}(\mathbf{r}) + V_H(\mathbf{r}) + V_{XC}(\mathbf{r})\right]\psi_i(\mathbf{r}) = \epsilon_i\psi_i(\mathbf{r}). \tag{3.43}$$

The first two potentials, in order, are the external potential and the same Hartree potential we have seen in equation (3.23). Recall that the latter describes the Coulomb repulsion between a single electron and the total electron density. The third potential, $V_{XC}(\mathbf{r})$, is the hitherto defined exchange-correlation potential describing such interactions. We will see that this potential is a conception that generalizes the Hartree formalism to contain both exchange and correlation effects.

In 1964, Kohn and Sham set out to extract the inherent Hartree equations from the formally exact Hohenberg-Kohn variational principle for the energy [58]. This is a delighting example of intuitive physics since the variational principle must not only contain the Hartree equations, but also residual information - corrections and improvements to Hartree formalism - on account of the fact that it is formally exact. In so doing, a Hartree-like formalism is deduced that is no harder to solve, but exact in nature.

Consider the variational principle for a system of N interacting electrons,

$$E[n(\mathbf{r})] = F[n(\mathbf{r})] + \int n(\mathbf{r}) V_{ext}(\mathbf{r}) d\mathbf{r} \geq E_0[n(\mathbf{r})], \tag{3.44}$$

in which the corresponding stationary condition is given by:

$$\frac{\delta}{\delta n(\mathbf{r})} \left[E[n(\mathbf{r})] - \mu \left(\int n(\mathbf{r}) d\mathbf{r} - N \right) \right] = 0, \tag{3.45}$$

where μ is the Lagrange multiplier that guarantees particle conservation. Accordingly, it is given by:

$$\frac{\delta E[n(\mathbf{r})]}{\delta n(\mathbf{r})} = \frac{\delta F[n(\mathbf{r})]}{\delta n(\mathbf{r})} + V_{ext}(\mathbf{r}) = \mu. \tag{3.46}$$

Recall that the exact universal functional, $F[n(\mathbf{r})]$, in equation (3.44) is unknown. At this juncture Kohn and Sham, therefore, *deliberately* partition it into the following three terms:

$$\begin{aligned} F[n(\mathbf{r})] &= T[n(\mathbf{r})] + V[n(\mathbf{r})] \\ &= T_s[n(\mathbf{r})] + E_H[n(\mathbf{r})] + E_{XC}[n(\mathbf{r})]. \end{aligned} \tag{3.47}$$

The first term is the kinetic energy functional for *non-interacting* electrons expressed in terms of the single-particle wavefunctions ψ_i,

$$T_s[n(\mathbf{r})] = -\frac{\hbar^2}{m}\sum_i^N \int \psi_i^*(\mathbf{r})\nabla^2\psi_i(\mathbf{r})d\mathbf{r}; \tag{3.48}$$

the second term is the classical interaction (Hartree) energy functional for pairs of electrons,

$$E_H[n(\mathbf{r})] = \frac{1}{2}\int\int \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}d\mathbf{r}d\mathbf{r}', \tag{3.49}$$

and the last term, $E_{XC}[n(\mathbf{r})]$, is the implicit definition for the exchange-correlation energy functional.

The concerning reason why $F[n(\mathbf{r})]$ is partitioned this way was primarily motivated by the fact that, in elementary density functional theory, the principal deficiency resulted from a rudimentary approximation for the exact kinetic energy $T[n(\mathbf{r})]$. Intending to produce better results, Kohn and Sham therefore recognized that if they decomposed the kinetic energy into two parts; the formally known kinetic energy functional for non-interacting electrons (3.48), and the kinetic energy functional for correlated electrons,

$$T_c[n(\mathbf{r})] = T[n(\mathbf{r})] - T_s[n(\mathbf{r})], \tag{3.50}$$

the error in approximating the exact kinetic energy in the elementary method can be reduced. In this way, notice that it *partly* allows for an exact treatment of the kinetic energy which, indeed, correspondingly reduces the total error. In keeping with this scheme, Kohn and Sham also split the exact potential energy $V[n(\mathbf{r})]$ into two parts: the well established Hartree energy functional, and the unknown exchange energy functional,

$$E_x[n(\mathbf{r})] = V[n(\mathbf{r})] - E_H[n(\mathbf{r})]. \tag{3.51}$$

This conjointly allows, at least to some degree, an exact treatment of the potential energy.

The question of why $F[n(\mathbf{r})]$ is partitioned in this way, however, still persists to some degree, and a more honed question ought to be asked: What is the utility of splitting the exact energy functionals, defined by equations (3.50) and (3.51)? The answer is two-fold. Firstly, the utility resides in the fact that doing so effectively aggregates the formally unknown terms into *one* unknown functional, $E_{XC}[n(\mathbf{r})]$. This allows for a theoretical formulation that can exclusively dispose all of its resources in developing better approximations for only one unknown implicitly defined term instead of widely

distributing its effort to account for exactly the difficult $T[n(\mathbf{r})]$ and $V[n(\mathbf{r})]$ terms. Secondly, and more importantly, the partitioned form for $F[n(\mathbf{r})]$ allows us to interpret equation (3.40),

$$E[n(\mathbf{r})] = F[n(\mathbf{r})] + \int n(\mathbf{r}) V_{ext}(\mathbf{r}) d\mathbf{r},$$

as the energy functional of a *non-interacting* classical electron gas subject to an effective potential $V_{eff}(\mathbf{r})$. To that end, given equation (3.47), rewrite equation (3.46) as:

$$\frac{\delta T_s[n(\mathbf{r})]}{\delta n(\mathbf{r})} + V_{eff}(\mathbf{r}) = \mu, \tag{3.52}$$

where

$$V_{eff}(\mathbf{r}) = V_{ext}(\mathbf{r}) + V_H(\mathbf{r}) + V_{XC}(\mathbf{r}), \tag{3.53}$$

and the Hartree potential and the exchange-correlation potential are respectively given by,

$$V_H(\mathbf{r}) = \frac{\delta E_H[n(\mathbf{r})]}{\delta n(\mathbf{r})} = \int \frac{n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' \tag{3.54}$$

and

$$V_{XC}(\mathbf{r}) = \frac{\delta E_{XC}[n(\mathbf{r})]}{\delta n(\mathbf{r})}. \tag{3.55}$$

Accordingly, Kohn and Sham showed that in actually considering a system that contained *non-interacting* electrons subject to the same effective potential (3.53), the resulting Euler-Lagrange equation would be exactly given by that for the case of *interacting* electrons (3.52) [59]. Therefore, since in the non-interacting case, the electronic density $n'(\mathbf{r})$ is obtained by calculating the wavefunctions $\psi_i'(\mathbf{r})$ of the single particle Schrödinger equation, we conclude by reason of a logical extension that we can also obtain the electron density $n(\mathbf{r})$,

$$n(\mathbf{r}) = \sum_{i=1}^{N} |\psi_i(\mathbf{r})|^2, \tag{3.56}$$

for the interacting case by solving the corresponding single-particle Schrödinger equation,

$$\left[-\frac{\hbar^2}{2m}\nabla^2 + V_{ext}(\mathbf{r}) + V_H(\mathbf{r}) + V_{XC}(\mathbf{r})\right]\psi_i(\mathbf{r}) = \epsilon_i\psi_i(\mathbf{r}).$$

Therefore, the partitioned form for $F[n(\mathbf{r})]$ allows for a convenient and familiar method to find the ground-state density without explicitly making use of the second theorem of Hohenberg and Kohn. These equations, which are formally known as the Kohn-Sham equations, are those given in the opening of this section (3.43).

Of significant interest is the knowledge that the Kohn-Sham equations are the exact categorical formulation of Hartree theory. Notice that if we appropriately neglect in equation (3.47) the exchange-correlation energy functional $E_{XC}[n(\mathbf{r})]$ (in which $V_{XC}(\mathbf{r})$ is implicit), the Kohn-Sham equations reduce to the Hartree equations (3.32). Therefore, since the Hartree equations must be solved self-consistently, it is not surprising that so too must the Kohn-Sham equations. To be formally rigorous, however, recognize that given the relations (3.54) and (3.55), the effective potential $V_{eff}(\mathbf{r})$ in the Kohn-Sham equations implicitly depends on the density. This has the distinctive meaning of necessarily self-consistent treatment of the corresponding equations.

In the Kohn-Sham method, as the complexity of a given system increases due to increasing N, the fundamental problem becomes no more difficult to solve. Consequently, since the Hartree (-Fock) method does not draw a parallel, the continued solvability of the increasingly complicated problem embodies the great achievement of the Kohn-Sham method, and by extension, DFT.

3.4 The Exchange-Correlation Functional

Although DFT is in principle an *exact* theoretical structure for viewing the electronic configuration of matter, in practice it is *inexactly* approximated because of the unknown exchange-correlation functional,

$$E_{XC}[n(\mathbf{r})] = T[n(\mathbf{r})] - T_s[n(\mathbf{r})] + V[n(\mathbf{r})] - E_H[n(\mathbf{r})]. \tag{3.57}$$

Recall that this functional is implicitly given by (3.47) and the defining equations

$$T_c[n(\mathbf{r})] = T[n(\mathbf{r})] - T_s[n(\mathbf{r})]$$

and

$$E_x[n(\mathbf{r})] = V[n(\mathbf{r})] - E_H[n(\mathbf{r})].$$

Therefore, the usefulness and practicality of DFT exclusively depends on the approximations used for $E_{XC}[n(\mathbf{r})]$.

The elemental point of departure for approximating the exchange-correlation functional is to divide it into two parts: the exchange part and the correlation part [60]. Accordingly, equation (3.57) is decomposed as:

$$E_{XC}[n(\mathbf{r})] = E_X[n(\mathbf{r})] + E_C[n(\mathbf{r})], \tag{3.58}$$

where $T_c[n(\mathbf{r})]$ contributes to $E_c[n(\mathbf{r})]$. The exchange part is known exactly and is defined by:

$$E_X^{exact}[n(\mathbf{r})] = -\frac{1}{2}\sum_{i,j}\int d\mathbf{r}d\mathbf{r}'\frac{\psi_i^*(\mathbf{r})\psi_i(\mathbf{r}')\psi_j^*(\mathbf{r}')\psi_j(\mathbf{r})}{|\mathbf{r}-\mathbf{r}'|}. \tag{3.59}$$

Notice that this definition is that for the exchange energy in Hartree-Fock theory - see the last term in equation (3.26) - with the exception that the Kohn-Sham wavefunctions serve as substitutions for those of Hartree-Fock. There is no known explicit expression for the correlation energy functional, and in consequence, it must be approximated.

At first glance, it appears that we have made progress in expressing $E_{XC}[n(\mathbf{r})]$ on account of the fact that instead of approximating the entire energy functional - which is seemingly more difficult - we need to only approximate a small unknown part, $E_C[n(\mathbf{r})]$. It turns out, however, that when the exact exchange functional is used in conjunction with an approximated correlation functional, it yields insufficiently accurate results for most properties of interest. The reason [60] behind this failure remained a scientific curiosity for a long time, nevertheless, it became clear that in order to marry these functionals in accordance with equation (3.58), a highly sophisticated approximation for the correlation functional was absolutely required. This proved to be difficult, and to that end, it can be shown that a simpler alternative is to give up the exact exchange functional, and in place, use an approximation. In this respect, an explicit error is brought about into $E_X[n(\mathbf{r})]$ which tends to cancel out the error in $E_C[n(\mathbf{r})]$. Therefore, using approximations for both functionals fortuitously results in successful approximations for $E_{XC}[n(\mathbf{r})]$ as a whole.

3.4.1 Local Density Approximation

The early thinking that prompted a reasonable implementation of DFT was founded upon a non-empirical and tentative idea [57]: obtain a nearly exact exchange-correlation functional for a trivial model and hope that the same result can be used for a more complicated system. To that end, the simplest model for which near exact results could be obtained is the homogeneous electron gas - a group of many interacting electrons moving in a constant external potential.

This model was studied by Thomas and Fermi in the early 1920s [61, 62]. There, it was shown that the non-interacting kinetic energy per unit volume is given by:

$$t_s^{hom}(n) = \frac{3\hbar^2}{10m}(3\pi^2)^{2/3}n^{5/3}, \tag{3.60}$$

where the electron density, n, is by definition a constant. However, in keeping with the above discussion we heuristically approximate locally for the *inhomogeneous* system:

$$t_s(\mathbf{r}) \approx t_s^{hom}[n(\mathbf{r})] = \frac{3\hbar^2}{10m}(3\pi^2)^{2/3}n(\mathbf{r})^{5/3}, \tag{3.61}$$

where in place n is the true electronic density $n(\mathbf{r})$. The comparative full kinetic energy is obtained by integrating over all space,

$$T_s[n(\mathbf{r})] = \int t_s^{hom}[n(\mathbf{r})]d\mathbf{r} = \frac{3\hbar^2}{10m}(3\pi)^{2/3}\int n(\mathbf{r})^{5/3}d\mathbf{r}, \tag{3.62}$$

since equation (3.61) is given per unit volume. This is a local density approximation (LDA) for the kinetic energy - though a seriously inadequate one [58]. Recall that this poor approximation used for the kinetic energy largely motivated Kohn and Sham to reshuffle the total energy functional $E[n(\mathbf{r})]$ in such a way that this error is to some degree redressed.

The local density approximation - though not good for imitating the non-interacting kinetic energy - proved to be very fruitful for approximating the exchange-correlation energy $E_{XC}[n(\mathbf{r})]$. Given that the exchange energy per volume is known exactly for a homogeneous system [63],

$$e_X^{hom}(n) = -\frac{3q^2}{4}\left(\frac{3}{\pi}\right)^{1/3}n^{4/3}, \tag{3.63}$$

(where q is the momentum transfer, a characteristic and dimensionless quantity), the corresponding LDA is:

$$E_x^{LDA}[n(\mathbf{r})] = -\frac{3q^2}{4}\left(\frac{3}{\pi}\right)^{1/3}\int n(\mathbf{r})^{4/3}d\mathbf{r}. \tag{3.64}$$

Notice that the constant density is similarly interchanged with $n(\mathbf{r})$ and $e_x^{hom}(n)$ integrated over all space to yield the full exchange energy. Be that as it may, the LDA for the correlation energy is a little more difficult since the corresponding homogeneous energy per unit volume, $e_C^{hom}(n)$, is not known exactly. However, such systems have been parametrized in the form of analytic functionals such as [64]:

$$\begin{aligned} E_C^{LDA}[n(\mathbf{r})] = -A\int n(a+\alpha_1 r_s) \\ \times\left[1+\frac{1}{A(\beta_1 r_s^{1/2}+\beta_2 r_s+\beta_3 r_s^{3/2}+\beta_4 r_s^2}\right]d\mathbf{r} \end{aligned} \tag{3.65}$$

where $r_s = (3/4\pi)^{1/3}$ and A, α_1, $\beta_1, \beta_2, \beta_3$, and β_4 are fixed parameters, and are therefore, known numerically.

Accordingly, the exchange-correlation LDA is commonly cast in the form of an integral expression:

$$E_{XC}[n(\mathbf{r})] = \int e_{XC}[n(\mathbf{r})]d\mathbf{r}, \tag{3.66}$$

where

$$e_{XC} = e_X^{hom} + e_C^{hom} \tag{3.67}$$

is the total exchange-correlation energy density - it has units of energy per unit volume. At times, however, a per-particle instead of per-volume notation is used, and as consequence, an additional $n(\mathbf{r})$ is included under the integral in equation (3.66). The corresponding exchange-correlation potential given by equation (3.55) is therefore written as:

$$V_{XC}(\mathbf{r}) = \frac{\delta E_{XC}[n(\mathbf{r})]}{\delta n(\mathbf{r})} = \left.\frac{\partial e_{XC}^{hom}(n)}{\partial n}\right|_{n=n(\mathbf{r})}. \tag{3.68}$$

Notice that in view of the Kohn-Sham equations (3.43) we now have all the information to formally extract the single-particle wavefunctions $\psi_i(\mathbf{r})$. We will elaborate this point in great detail below.

The LDA, although nearly exact for a homogeneous electron gas, was regarded *a priori* to be profitable for slowly varying densities exclusively[58]. However, in atomic systems of interest the density is generally rapidly varying and, therefore, the LDA appears to be a seriously crude approximation. Despite this attention, it gives extremely useful results (for many solid-state purposes) and to that end, is a reasonable accurate approximation to the true exchange-correlation energy [60]. It should not be surprising, however, that the LDA is to no degree the only attempt to approximate the exchange-correlation functional.

3.4.2 Generalized Gradient Approximation

Real systems have spatially varying electron densities, $n(\mathbf{r})$, and are consequently spatially inhomogeneous. It is thus serviceable to include information on the rate of the variations of the electron density in the functional. Therefore, open to the view of an exchange-correlation approximation that more faithfully represents nature, Hohenberg and Kohn proposed a systematic extension to the LDA known as the Gradient Expansion Approximation (GEA) [57]. It can be constructed as [60]:

$$E_{XC}[n] = \int e_{XC}^{LDA}(n)[1 + \mu(n)s^2 + ...]d\mathbf{r}, \tag{3.69}$$

where s is a dimensionless quantity known as the reduced density gradient,

$$s = \frac{|\nabla n|}{2k_F n}, \tag{3.70}$$

and is a measure of how fast and how much the density varies - a measure of homogeneity. As a companion to s is $\mu(n)$; the gradient coefficient. It is a function of the density that accordingly reduces to a constant. These coefficients (of low order gradient terms) can be determined non-empirically, and therefore, it is expected by reason of experience and intuition to assume that equation (3.69) is more accurate than the LDA. However, when the truncated GEA is implemented, it is found to be completely inadequate for the description of atoms and solids [65].

The Exchange-Correlation Hole

In order to understand why the gradient expansion is exceedingly unsuccessful, the concept of the exchange-correlation hole, $n_{XC}(\mathbf{r}, \mathbf{r}')$, must be established. It is a quantum-mechanical zone that surrounds and follows every electron in an interacting system. It has the effect of reducing the probability of finding other electrons within its immediate vicinity, and therefore, reflects the combined effect of the Pauli exclusion principle and the Coulombic electron-electron interaction. The physical hole is defined as [65]:

$$n_{XC}(\mathbf{r}, \mathbf{r}') = g(\mathbf{r}, \mathbf{r}') - n(\mathbf{r}'), \tag{3.71}$$

where $n(\mathbf{r}')$ is the average density of electrons at $\mathbf{r}'$, such that

$$\int n(\mathbf{r}')d\mathbf{r}' = N, \tag{3.72}$$

and $g(\mathbf{r}, \mathbf{r}')$ is the average density of electrons at the point $\mathbf{r}'$ given that there is one at $\mathbf{r}$. Upon that distinction, then

$$\int g(\mathbf{r}, \mathbf{r}')d\mathbf{r}' = N - 1. \tag{3.73}$$

Consequently, the hole satisfies a relevant normalization condition known as the sum rule,

$$\int n_{XC}(\mathbf{r}, \mathbf{r}')d\mathbf{r}' = -1. \tag{3.74}$$

It implies that if an electron is specifically localized at $\mathbf{r}$ then it is not present in the remaining system, and by extension, it asserts that an electron cannot interact with itself - which is a physical constraint. Concerning that matter, any derived density functional - to which a model hole is implied - must conform to equation (3.74) and other known properties of the exact hole [65]. Otherwise, extensive errors in the energy ensues by virtue of the *exact* expression for the exchange-correlation energy:

$$E_{XC}[n(\mathbf{r})] = \frac{1}{2}\int n(\mathbf{r})d\mathbf{r}\int \frac{n_{XC}(\mathbf{r}, \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}d\mathbf{r}'. \tag{3.75}$$

Notice that n_{XC} is the (coupling-constant-averaged) exchange-correlation hole, and it is therefrom readily apparent that a poor approximation to the hole directly leads to a poor estimate of the energy.

With that, Perdew and co-workers revealed that the exchange-correlation hole underlying the second-order gradient expansion (GEA) is not the hole of any physical system, and subsequently violates equation (3.74) and other stringent conditions. This incontestably leads to a poor approximation to the exchange-correlation energy and is the basis of the disappointing results. By the same argument, however, since the LDA of Section 1.4.1 adversely describes both the shape and size of the exchange-correlation hole remarkably well, it is admittedly reliable and extremely useful when implemented [66]. In any case, Perdew, Becke and others sequentially devised an elaborate procedure to correct the palpable faults of the gradient expansion approximation. Accordingly, they removed all the exhibiting unphysical contributions to the GEA hole and replaced the truncated series in equation (3.69) with an enhancement factor, $F_{XC}(n, s)$, such that it restores equation (3.74) and other physical constraints. In other words, the description of the average hole is *enhanced*. This in turn gives rise to density functional approximations of the form:

$$E_{XC}[n] = \int e_{XC}^{LDA}(n) F_{XC}(n, s) d\mathbf{r}, \tag{3.76}$$

where n is the electronic density and s is the reduced density gradient given by equation (3.70). Equations of this structure are called Generalized Gradient Approximations (GGA) and since the analytic form of the enhancement factor is separate and distinct for each case, many GGA functionals exist.

An important GGA functional is that of Perdew and Wang (PW91) which has been developed non-empirically, and hence embodies exact quantum-mechanical relationships [67]. By way of explanation, the PW91 functional is the first analytical fit to the second-order density-gradient expansion for the exchange-correlation hole of a given slowly varying density system. Of this process, the functional is subject to comply with the sum rule (3.74) and is sensitive to other exact constraints such that it approximately incorporates inhomogeneity effects while keeping the leading characteristics of the LDA [67]. In point of fact, the GGA functionals generally improves upon the LDA's account of atoms, molecules and solids.

3.4.3 Discussion

A significant characterization and the understanding of the physics of atoms, molecules and solids follow from Coulomb's law and the understanding of

quantum mechanics. Their underlying principles are universally recognized as cogent and well grounded in experiment. Starting from these principles, the computationally tractable structure of the Kohn-Sham density functional theory is developed and it can be shown that the ground-state exchange-correlation energy is a functional of the electronic density $n(\mathbf{r})$ [57]. The central objective and fundamental mathematical problem of DFT is exclusively defined once the exchange-correlation functional has been specified. See equations (3.43) and recall that

$$V_{XC}(\mathbf{r}) = \frac{\delta E_{XC}[n(\mathbf{r})]}{\delta n(\mathbf{r})}.$$

However, since the constituting E_{XC} as a whole is not known exactly, apparent and clever approximations are required. We have seen in the previous Sections that these approximations are commonly expressed in the from of integral expressions [60],

$$E_{XC}[n(\mathbf{r})] = \int e_{XC}(n, ...)d\mathbf{r}. \tag{3.77}$$

where the energy density, e_{XC}, is a function of density-dependent elements - such as the modulus of the gradient of the density, $|\nabla n|$, implicit in the GGA. Other common elements of e_{XC} are the Laplacian of the density, $\nabla^2 n$, or more popular is the kinetic energy density of the Kohn-Sham orbitals,

$$\tau(\mathbf{r}) = \frac{1}{2}\sum_{i}^{occ} |\nabla\psi_i(\mathbf{r})|^2, \tag{3.78}$$

and the exact Hartree-Fock exchange energy which is given by equation (3.59):

$$E_X^{exact}[n(\mathbf{r})] = -\frac{1}{2}\sum_{ij}\int d\mathbf{r}d\mathbf{r}' \frac{\psi_i^*(\mathbf{r})\psi_i(\mathbf{r}')\psi_j^*(\mathbf{r}')\psi_j(\mathbf{r})}{|\mathbf{r}-\mathbf{r}'|}.$$

These functionals are classified as a systematic non-empirical approach to density functional approximations. The alleged "constraint satisfaction" method accepts the known mathematical properties of the exact exchange-correlation functional as *constraints*, and correspondingly constructs a functional that satisfies those constraints. In this approach, the more explicit properties the approximations have in accordance with the exact functional,

the more accurate and universal they will be. It is argued by Perdew and co-workers that this approach is the most convincing and enduring one [68].

Chapter 4

Implementing Density Functional Theory

Of the various excitations of a crystalline solid, the most important are lattice waves and electron states. To bring these into effect, earliest parts of the theory assumed that the ionic motion of the electrons could be separated and later pieced together. Fortunately, this treatment is not spurious and it is well grounded (on the account of Born and Oppenheimer) that each excitation makes a separate contribution to the total energy in a qualitative and reliable way [52]. From this sprung a culmination of ornate pinnacles of calculations and interpretations of solids [12]. The line of these achievements are commendable and deserve many words. However, this would go well outside the limits of this work and on another note, life is short. Instead, the aim here has been to make clear the physical model, the fundamental problem, and subsume the perfect crystalline solid to mathematics.

The *physical model* is understood in principle. Imagine a free neutral atom. It consists of a nucleus enclosed by tightly bound electrons that when considered together forms a halo of charge, known as the core. About this core are one or more valence electrons whose orbital motion is impelled by the permeating and fluctuating electric fields. When the atoms are collected together to form a solid, the orbitals overlap and strongly interact through the Coulomb potential and the quantization scheme in which each valence electron is localized on its atom breaks down. The electrons instead tear through the array of periodic ions as though they barely exist and fill pools of states of the solid as a whole. Attached to this scene are atom vibrations that compress and expand the lattice. This affects the electrostatic potential

acting on the free electrons and occasionally causes them to be deflected out of their customary state.

The picture has been set forth and the *fundamental problem* is the portrayal of its structure and dynamics. Unfortunately, the totality of every coordinate of every particle is mathematically fraught and so complicated that the scientist, who knows the arguments upon which the theory is founded, must appeal to physical plausibility. "It becomes desirable that approximate practical methods of applying quantum mechanics should be developed, which can lead to explanation of the main features of complex atomic systems without too much computation." To that end, density functional theory is among the most practical and accurate schemes to procure information about the dynamics and structure of the many-body system. While it is true that DFT can predict the frequencies of lattice vibrations, it is not yet available as a routine option in the most widely used packages. Therefore, DFT is mainly concerned with the dynamics and structure of the many *electron* problem alone.

4.1 Planewave Implementation

Atoms in an idealized infinite crystal are organized in complicated periodic arrays and strongly interact through the Coulomb force and the so-called exchange force. It is, therefore, colloquial to first examine the simpler problem of non-interacting electrons in a static and periodic potential. This matter, not lacking in subtlety, is well-considered by Bloch's theorem [12] which asserts the existence of a good quantum number $\mathbf{k}$ such that

$$\varphi_{\mathbf{k}}(\mathbf{r}+\mathbf{l}) = e^{i\mathbf{k}\cdot\mathbf{l}}\varphi_{\mathbf{k}}(\mathbf{r}), \tag{4.1}$$

where $\mathbf{l}$ is a direct lattice vector. The most general solution to this boundary condition is

$$\varphi_{\mathbf{k}} = e^{i\mathbf{k}\cdot\mathbf{r}}\sum_{\mathbf{G}} c_{\mathbf{G}}(\mathbf{k})e^{i\mathbf{G}\cdot\mathbf{r}} \tag{4.2}$$

$$= e^{i\mathbf{k}\cdot\mathbf{r}}u_{\mathbf{k}}(\mathbf{r}) \tag{4.3}$$

where $\mathbf{G}$'s are reciprocal lattice vectors and $\mathbf{u_k}(\mathbf{r})$ is a necessarily periodic function of the ionic lattice,

$$u_{\mathbf{k}}(\mathbf{r}+\mathbf{l}) = u_{\mathbf{k}}(\mathbf{r}). \tag{4.4}$$

The principal fact of the crystal is that $u_{\mathbf{k}}(\mathbf{r})$ is the potential of the lattice [12] and as it appears, a trivial application of Fourier's theorem,

$$u_{\mathbf{k}}(\mathbf{r}) = \sum_{\mathbf{G}} u_{\mathbf{G}}(\mathbf{k}) e^{i\mathbf{G}\cdot\mathbf{r}}, \tag{4.5}$$

reproduces Equation (4.2) upon substitution.

Of fundamental importance in the theory of solids is the inferred *free* electron approximation - the colloquial simpler problem of a constant potential $u_{\mathbf{k}}(\mathbf{r})$. In this situation, the electrons automatically fall into planewave states,

$$\varphi_{\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}} \tag{4.6}$$

with band energies

$$\epsilon(\mathbf{k}) = \frac{\hbar^2\mathbf{k}^2}{2m}. \tag{4.7}$$

Given its simplicity, it is surprisingly successful in explaining many experimental phenomena [12]. However, with the increasing understanding of solids it is important to consider the effects of the ionic lattice, and to that end, it is deserving to let a small periodic perturbing potential act on the free electrons.

The wavefunctions will no longer be pure planewaves, (4.6), but mixtures. To first order, the perturbed state is [12],

$$\varphi_{\mathbf{k}}(\mathbf{r}) = \phi_{\mathbf{k}} + \sum_{\mathbf{G}} \frac{u_{\mathbf{G}}\phi_{\mathbf{k}+\mathbf{G}}}{\hbar^2[\mathbf{k}^2 - (\mathbf{k}+\mathbf{G})^2]/2m}, \tag{4.8}$$

where $\phi_{\mathbf{k}}$ is the initial free electron state. The corresponding energy is

$$\epsilon(\mathbf{k}) = \frac{\hbar^2 k^2}{2m} + u_0 + \sum_{\mathbf{G}} \frac{|u_{\mathbf{G}}|^2}{\hbar^2[\mathbf{k}^2 - (\mathbf{k}+\mathbf{G})]^2/2m}, \tag{4.9}$$

and causes wonder. How good is this expansion? The energy denominators that occur in (4.9) indicate that the solution is not valid when

$$|\mathbf{k}| = |\mathbf{k} - \mathbf{G}|. \tag{4.10}$$

Accordingly, the denominator vanishes and Equation (4.9) blows up. In this event, the states are degenerate in energy; the corrections are no longer small and it becomes necessary to apply the "appropriate" method. However, prior to such a consideration, the geometrical identification of $\mathbf{k}$ in Equation (4.10) is essential.

The tip of the corresponding $\mathbf{k}$ vector lies on the perpendicular bisector of the reciprocal lattice vector, $\mathbf{G}$, which dissect reciprocal space into distinct regions known as Brillouin zones. The central region, known as the first Brillouin zone, is of special importance and has special properties that will become readily apparent in a practical DFT calculation. Of general importance is the statement that each Brillouin zone construction gives a vivid geometrical interpretation of the diffraction condition [12]. That is to say, any electron, $\varphi_{\mathbf{k}}$, whose wavevector terminates on a zone boundary will satisfy the diffraction condition and, as it happens, will be Bragg reflected. This introduces energy gaps in the band structure of the solid and are of considerable importance in determining whether a solid is an insulator or conductor.

The fibrous and basic *mathematical* question strictly remains: How do we find a solution in the neighbourhood of the zone boundary (4.10)? There, it becomes necessary to consider the perturbation equations explicitly. On that, the matrix elements of the perturbing potential between different unperturbed states of the same energy are,

$$u_{\mathbf{k}\mathbf{k}'}(\mathbf{r}) = \int \varphi^*_{\mathbf{k}}(\mathbf{r}) u_{\mathbf{k}}(\mathbf{r}) \varphi_{\mathbf{k}'}(\mathbf{r}) d\mathbf{r}, \tag{4.11}$$

where $\varphi_{\mathbf{k}}$ and $\varphi_{\mathbf{k}'}$ are plane-wave states (4.6). However, the symmetry of the lattice implies that

$$u_{\mathbf{k}\mathbf{k}'}(\mathbf{r}) = \begin{cases} u_{\mathbf{G}} & \text{if } \mathbf{k} + \mathbf{G} - \mathbf{k}' = 0 \\ 0 & \text{otherwise} \end{cases}. \tag{4.12}$$

This means that the perturbing potential only connects states that are separated by a reciprocal lattice vector, and as a result, the wave-function can be expanded as

$$\varphi_{\mathbf{k}}(\mathbf{r}) = \sum_{\mathbf{G}=0} \alpha_{\mathbf{k}-\mathbf{G}} e^{i(\mathbf{k}-\mathbf{G})\cdot\mathbf{r}}. \tag{4.13}$$

Upon substituting into the Schrodinger equation and multiplying through by one term in the corresponding series, it can be shown that

$$\left\{\epsilon^0_{\mathbf{k}-\mathbf{G}} - \epsilon(\mathbf{k})\right\}\alpha_{\mathbf{k}-\mathbf{G}} + \sum_{\mathbf{G}'} u_{\mathbf{G}'-\mathbf{G}}\alpha_{\mathbf{k}-\mathbf{G}'} = 0. \tag{4.14}$$

Although this is a useful form of the wave equation in a periodic potential, it is quite formidable because there are, in principle, an infinite number of coefficients $\alpha_{\mathbf{k}-\mathbf{G}'}$ to be determined. As an approximation, it is insightful to ignore all but only two in (4.14). That is,

$$\left.\begin{aligned}\left\{\epsilon^0_{\mathbf{k}} - \epsilon(\mathbf{k})\right\}\alpha_{\mathbf{k}} + u_{\mathbf{G}}\alpha_{\mathbf{k}-\mathbf{G}} = 0 \\ u_{-\mathbf{G}}\alpha_{\mathbf{k}} + \left\{\epsilon^0_{\mathbf{k}-\mathbf{G}} - \epsilon(\mathbf{k})\right\}\alpha_{\mathbf{k}-\mathbf{G}} = 0.\end{aligned}\right\} \tag{4.15}$$

The corresponding secular equation is quadratic in $\epsilon(\mathbf{k})$, with solutions:

$$\epsilon(\mathbf{k}) = \frac{1}{2}\left\{\frac{\hbar^2\mathbf{k}^2}{2m} + \frac{\hbar^2(\mathbf{k}+\mathbf{G})^2}{2m} \pm \left[\left(\frac{\hbar^2\mathbf{k}}{2m} - \frac{\hbar^2(\mathbf{k}+\mathbf{G})^2}{2m}\right)^2 + 4|u_{\mathbf{G}}|^2\right]^{\frac{1}{2}}\right\}. \tag{4.16}$$

The two roots give the dominant effect of the periodic potential on the free electron states $e^{i\mathbf{k}\cdot\mathbf{r}}$ and $e^{i(\mathbf{k}-\mathbf{G})\cdot\mathbf{r}}$ of Equation (4.15). From this, the concerning breakdown of Equation (4.9) at the zone boundary is clearly understood: For any values of $\mathbf{k}$ that satisfy the Bragg condition, Equation (4.16) becomes

$$\epsilon(\mathbf{k}) = \epsilon^0_{\mathbf{k}} \pm |u_{\mathbf{G}}| \tag{4.17}$$

and one level is raised in energy by $|u_{\mathbf{G}}|$ while the other is decreased by the same amount. This means that the effect of the perturbation is to introduce jump discontinuities into the energy function and is ultimately related to the theory of Bragg diffraction, which itself, is regarded as synonymous with Equation (4.10).

4.1.1 Planewaves

Calculating $\epsilon(\mathbf{k})$ is of fundamental importance in the theory of solids and is regarded as a specialized art. The delineating details of the previous discussion, however, inferred a rigorous method. To begin, determine the coefficients from (4.14) and expand the wavefunction accordingly. Retain only the terms that can be mixed by the perturbation close to the zone boundary and solve the secular equation (4.15) for $\epsilon(\mathbf{k})$.

Figure 4.1: The electronic wave function of an electron in a crystal.

This is known as the *nearly free electron* model and is useful as a means to illustrate the quantitative behaviour of energy bands in reciprocal space. It is, however, just that - not practically useful. The electric field near the core of an ion is very strong and decreases as $1/G^2$ [12]. Therefore, the Fourier components of the potential,

$$u_{\mathbf{k}}(\mathbf{r}) = \sum_{\mathbf{G}} u_{\mathbf{G}}(\mathbf{k}) e^{i\mathbf{G}\cdot\mathbf{r}},$$

should quickly tend to zero with increasing magnitude of $\mathbf{G}$. Unfortunately, this is not the case and ensues many terms in the wavefunction expansion (4.13). The method is thus only suited to the description of the lowest few bands where a large number of planewaves is not necessary.

Does this mean that a planewave basis set cannot be used? Not necessarily. Look at the wave function for an electron in a lattice (Figure (4.1). It essentially consist of a nearly perfect sinusoidal wave in the interstitial region of the ions and a rapidly varying tail near the atomic nuclei. These characteristic oscillations approximate atomic orbitals - requiring a very high order of Fourier components - and express the fact that electrons near the core are tightly bound. From a physical standpoint, however, core electrons are not especially important in the chemistry of the solid. Chemical bonding and other physical properties are dominated by the less tightly bound valence electrons. In that, it became clear from the earliest development of planewave methods that there could be great advantages by approximating the properties of core electrons in a way that could reduce the number of Fourier components in the calculation.

It is quite common in applications to invoke the *frozen core* approximation, in which the lowest-lying electrons are constrained in an atomic reference configuration. This allows for the description of a "pseudo"-wavefunction with far fewer Fourier components, thus employing planewaves as a practical

basis set. It is with good intention to make mention of the fact that although the core electrons are strongly bound in most cases, the approximation breaks down for elements with extended core states (also known as semi-core states) and require special considerations:- to be addressed in Section (4.2.3).

4.1.2 Pseudopotentials

A function such as that shown in Figure (4.1) can be generated in a numerous of mathematical ways. A particularly significant and clever approach is to subtract from a planewave, φ_{PW}, a Bloch sum formed from lower lying core states, $\varphi_c(\mathbf{r})$ [69]. That is,

$$\phi_{\mathbf{k}} = \varphi_{PW} - \sum_c \langle \varphi_c | \varphi_{PW} \rangle \, \varphi_c, \tag{4.18}$$

where the summation is over all the core states. This works because the core states,

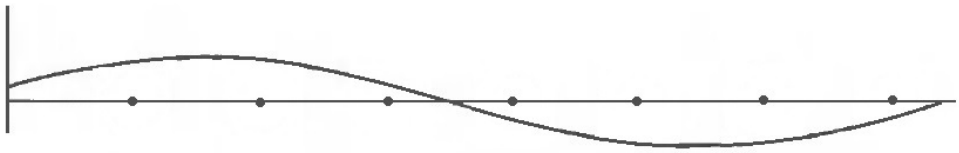
$$\varphi_c(\mathbf{r}) = \sum_{\mathbf{l}} e^{i\mathbf{k}\cdot\mathbf{l}} \varphi_c(\mathbf{r} - \mathbf{l}), \tag{4.19}$$

written as a *linear combination of atomic orbitals*, are strongly localized. Each atomic orbital corresponds to the probabilistic localization of a single electron on a given atom, l, in any cell of the crystal. This way, Equation (4.19) - like any Bloch sum - gives a good (local) solution to the Schrodinger equation of the entire crystal, as seen in Figure (4.2b).

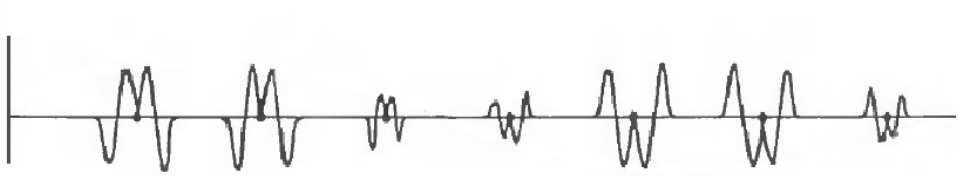
A crystal, however, does not only contain states making up the ion core. There is a sea of higher states that extend throughout the entire crystal and are certainly solutions to the same Schrodinger equation. It therefore becomes by necessity that the higher states be orthogonal to the core states,

$$\langle \phi_{\mathbf{k}} | \varphi_{c\mathbf{k}} \rangle = 0. \tag{4.20}$$

Coupled with this and the fact that the higher states must be as smoothly varying as a planewave in the region between the atoms, an appropriate linear combination can be written down at once - Equation (4.18). By explicit construction, $\phi_{\mathbf{k}}$ looks like a planewave in the interstitial region, where each φ_c are highly localized, but within the core, it is orthogonal to all the core states [16] and will therefore have all the required rapid oscillations.



(a) Planewave



(b) Core function

(c) $\phi_{\mathbf{k}}$ = planewave - core function

Figure 4.2: The synthesis of an electronic wavefunction in a crystal [16].

Equation (4.18) is a good guess; it nicely represents a state running through a crystal (see Figure (4.2)). Herewith the fact that it also satisfies the Bloch condition, one can use $\phi_{\mathbf{k}}$ as an *orthogonalized planewave* (OPW) basis set for the wavefunction as a solution of the Schrodinger equation. The expansion will be of the form:

$$\psi_{\mathbf{k}} = \sum_{\mathbf{G}} \alpha_{\mathbf{k}-\mathbf{G}} \phi_{\mathbf{k}-\mathbf{G}}, \tag{4.21}$$

where the coefficients can be determined by minimizing the expectation value of the energy, as demanded by the variational principle (see Appendix B). Equation (4.21) is similar to a Fourier expansion, only it is sufficiently modulated at the nuclei so that it converges rapidly. As it happens, in simple metals and semiconductors, only a few components are necessary to give an accurate approximation of a valence electron over a large region of a Brillouin zone.

As an extension to this method, one can substitute (4.18) into the Schrodinger equation $H\psi = \epsilon\psi$, to find [70]

$$H\varphi_{PW} + \sum_{c} (\epsilon - \epsilon_c) \langle \varphi_c | \varphi_{PW} \rangle \varphi_c = \epsilon \varphi_{PW}, \tag{4.22}$$

since φ_c is an eigenstate of H with energy ϵ_c, and

$$\varphi_{PW} = \sum_{\mathbf{G}} \alpha_{\mathbf{k}-\mathbf{G}} e^{i(\mathbf{k}-\mathbf{G})\cdot\mathbf{r}}. \tag{4.23}$$

If one defines, as a formal operator,

$$u_R \varphi_{PW} \equiv \sum_{c} (\epsilon - \epsilon_c) \langle \varphi_c | \varphi_{PW} \rangle \varphi_c, \tag{4.24}$$

then Equation (4.22) can be neatly written as:

$$(H + u_R)\varphi_{PW} = \epsilon \varphi_{PW}. \tag{4.25}$$

This is quite remarkable. It is as though the solution everywhere in space is a simple planewave in an *effective potential*. The Hamiltonian of the novel Schrodinger equation is

$$H + u_R = \frac{-\hbar^2}{2m}\nabla^2 + u + u_R, \tag{4.26}$$

from which one writes as the corresponding effective potential,

$$u_{eff} = u + u_R. \tag{4.27}$$

The formal tools of Section (4.1) may now be used. The coefficients $\alpha_{\mathbf{k}-\mathbf{G}}$ of Equation (4.23) satisfy equations like (4.14), but with Fourier components of u_{eff} instead of those of the true lattice potential u.

The contribution u_R to the true lattice potential is not exact. It turns out that u_R must be different when it operates on functions of different angular momentum,

$$u_R = u_s + u_p + u_d + \cdots , \tag{4.28}$$

where u_s only operates on functions with s-symmetry, etc. The process of constructing u_R is, therefore, not unique. This means that the wavefunctions in Equation (4.25) are, in fact, *pseudo*-wavefunctions $\varphi^{PS}(\mathbf{r})$; they depend on the construction of the pseudopotential u_R. This is the essence of the pseudopotential approximation: $u^{PS}(\mathbf{r})$ replaces the electron density of a selected set of core electrons with one designed to match important physical and mathematical properties of the true ion core. This produces a smooth tailed wavefunction inside a chosen core radius r_c (as seen in Figure (4.3)). In outer regions of the core, the pseudopotential gently transitions into the true potential of the crystal and the corresponding pseudo-wavefunction $\varphi^{PS}(\mathbf{r})$ imitates the physically important valence electrons. This way, only low $|\mathbf{G}|$ planewaves are required (which means the method is computationally efficient) and the complicated nature of the scheme is, in an appropriate manner, transferred from the calculation itself to the development of the pseudopotential. However, recall that there is no unique method to generate a pseudopotential.

4.2 Augmented Planewave Implementation

The overarching and recurring objection to band structure calculations has been the inability to give a detailed, distinctive, and efficient account of the wavefunction in a crystal. On the one hand, a simple linear combination of atomic orbitals cannot easily represent the nearly free electron waves - they fall off too rapidly in the region outside the core - while on the other hand, a simple planewave expansion (4.13) cannot behave like an atomic orbital

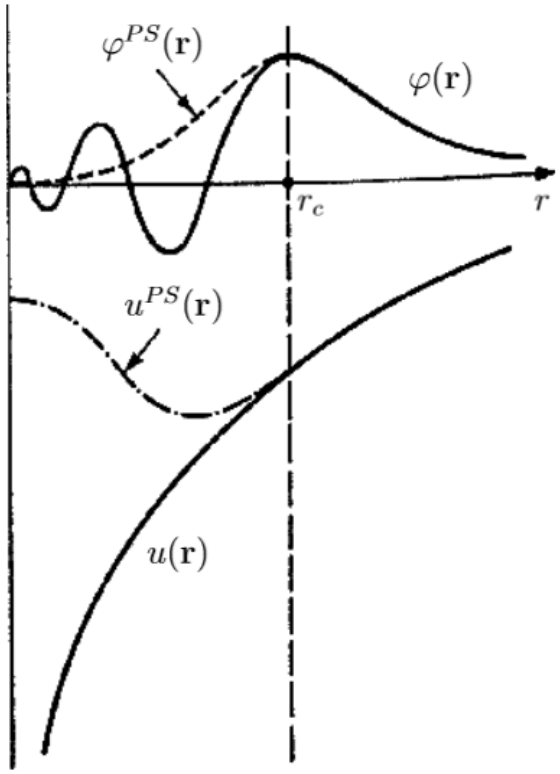


Figure 4.3: The replacement of the all-electron wavefunction and core potential by a pseudo-wavefunction and pseudoptential. Notice that outside the core region (defined by r_c), the pseudopotential becomes the true potential, and the pseudo-wavefunction smoothly transitions into the valence state wavefunction. [70]

inside the core of the ion; it would require too many terms. Given the mutually exclusive nature of the problem, one may be inspired to construct a Fourier-like expansion out of valence states orthogonalized to the lower lying core states, as in Equation (4.21), to set it right. This way, the corresponding wavefunction is both sufficiently modulated at the nuclei and smoothly varying in the interstitial region so that it converges rapidly.

This picture is, indeed, very likeable. It has been successfully applied - mostly in the simplified version of the pseudopotential method - to almost all types of solids and gives quite good approximations of the band structure. However, a function such as that shown in Figure (2.1) can be generated in a completely different way, and that is by looking at the potential function itself. To bring in and establish an additional scheme might seem to lack substance and worth, but it will become abundantly clear that it is valuable.

4.2.1 The APW Method

At an atomic site, the potential resembles that of a free atom; it is strongly varying but spherically symmetrical. In the interstices of the ion cores, however, the potential is very flat and constant. To wit, space can be divided into distinct regions: non-overlapping spheres about each atomic site and the remaining volume between the spheres (Figure (4.4)). This way, the Schrodinger equation can be solved exactly within each sphere, in spherical harmonics, and in the remaining interstitial region, in planewaves. That is,

$$\varphi(\mathbf{r}) = \begin{cases} \Omega^{-1/2} \sum_{\mathbf{G}} c_{\mathbf{G}} e^{i(\mathbf{G}+\mathbf{k})\cdot\mathbf{r}} & \mathbf{r} \in I \\ \\ \sum_{lm} A_{lm} u_l(r) Y_{lm}(\mathbf{r}) & \mathbf{r} \in S \end{cases} \tag{4.29}$$

where Ω is the cell volume, $c_{\mathbf{G}}$ and A_{lm} are expansion coefficients, $Y_{lm}(\mathbf{r})$ are spherical harmonics, and $u_l(r)$ are solutions to the radial part of the Schrodinger equation inside the sphere,

$$\left[-\frac{d^2}{dr^2} + \frac{l(l+1)}{r^2} + V(r) - E_l \right] r u_l(r) = 0. \tag{4.30}$$

Here, E_l is an undetermined parameter (set to the eigenvalue of the band energy) and V is the spherical component of the potential in the sphere.

The dual representation of $\varphi(\mathbf{r})$ in (4.29) is not guaranteed to exactly match over the entire surface of the spherical boundary (a constraint necessary for the kinetic energy to be well defined). It must be strictly imposed

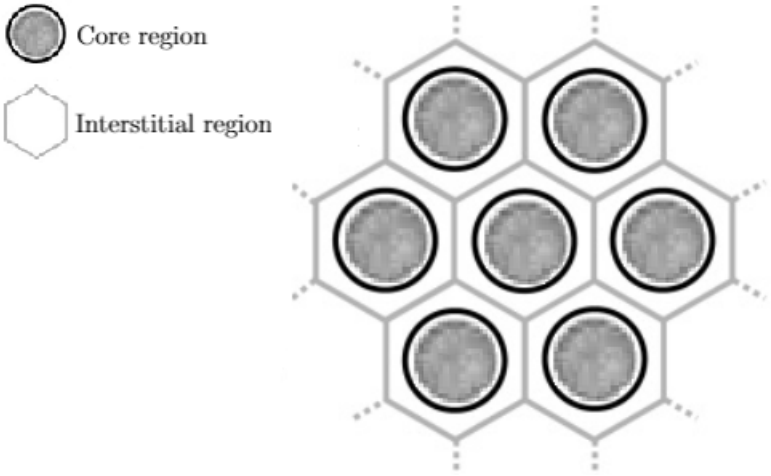


Figure 4.4: The partitioning of space into core and interstitial regions.

and the procedure is fairly straightforward: The solution of the Schrodinger equation within each sphere is of the aforementioned radial form,

$$\varphi_r(\mathbf{r}) = \sum_{lm} A_{lm} u_l(r) Y_{lm}(\mathbf{r}),$$

where the coefficients, A_{lm}, are *chosen* so that $\varphi(\mathbf{r})$ exactly matches at the spherical surface to a single planewave (expanded in spherical harmonics),

$$e^{i\mathbf{k}\cdot\mathbf{r}} = 4\pi \sum_{lm} i^l j_l(kr) Y^*_{lm}(\hat{\mathbf{k}}) Y_{lm}(\hat{\mathbf{r}}). \tag{4.31}$$

Indeed j_l are Bessel functions of order l, and the matching $\varphi_r(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}}$ transpires if

$$A_{lm} = \frac{4\pi i^l}{\sqrt{\Omega} u_l(R,E)} \sum_{\mathbf{G}} c_{\mathbf{G}} j_l(|\mathbf{G}+\mathbf{k}|R) Y^*_{lm}(\mathbf{G}+\mathbf{k}), \tag{4.32}$$

where the origin of the system ($R = 0$) is chosen to coincide with the nucleus of the atom.

With this, a single *augmented planewave* (APW) is defined. By putting (4.32) into (4.29), one can see that $\varphi(\mathbf{r})$ is an oscillatory function travelling through the crystal whose simple behaviour is changed into something more complicated inside the *muffin-tin* sphere of an encountered atom. The

advantage of this argument is that it clearly reflects the actual physics of the lattice: An electron, represented by a planewave, travels through the crystal and undergoes multiple scattering at the atoms. If for some energy E_l, the outgoing scattered waves interfere destructively, a bound state has been determined. Unfortunately, it turns out that this exposition is not proficient in the description of the electronic structure. The labour in fitting the continuity condition is very heavy and practically intractable for systems of modest complexity. In principle, an infinite number of energy dependent coefficients, A_{lm}, are required to create the matching. The major difficulty, however, resides not in the infinite sum, but in the implicit dependency of energy parameter E_l.

Notice that the augmented planewave is *not* a solution of the Schrodinger equation of the whole lattice. Indeed, there is no special relationship between E_l and $\mathbf{k}$ in the interstitial region. This means that the energy cannot be taken as a fixed parameter in the construction of the augmented planewave basis. Instead, it must be set to a variational estimate of the energy of the correct solution of the whole lattice: $E_l = \epsilon(\mathbf{k})$. The APW method, therefore, tries to approximate the solution of the crystal Schrodinger equation by a linear combination of augmented planewaves, all with the same energy.

From the Bloch theorem, the expansion must be of the form

$$\psi_{\mathbf{k}}(\mathbf{r}) = \sum_{\mathbf{G}} \alpha_{\mathbf{k}-\mathbf{G}} \varphi_{\mathbf{k}-\mathbf{G}}, \tag{4.33}$$

where $\alpha_{\mathbf{k}-\mathbf{G}}$ are variational coefficients. Progress has been made; by setting the energy of the APW to the actual band energy of the Bloch wavefunction, $\psi_{\mathbf{k}}(\mathbf{r})$ is guaranteed to satisfy the Schrodinger equation in both the atomic and interstitial regions. Unfortunately, the radial wavefunctions $u_l(r, E_l)$, making up each basis function, now depends implicitly on the band energy $E_l = \epsilon(\mathbf{k})$. As a consequence, the Hamiltonian (set up in terms of this basis) cannot be determined by a simple diagonalization and the corresponding solution to the secular equation becomes a nonlinear problem. Accordingly, the energy occurs explicitly in both the diagonal and off-diagonal matrix elements and, as a result, the only practical way of solving the secular equation is to evaluate its determinant

$$|H - \epsilon_i S| = 0, \tag{4.34}$$

as a function of energy and use an interpolation procedure to determine its

roots. This means that the energy bands cannot be obtained from a simple diagonalization and, therefore, a great deal of computing is required.

The band structure is very sensitive in detail to the form of the atomic potential, and this, in turn, depends on how it is defined. In the muffin-tin approximation (in which the potential is spherically symmetric inside the spheres and constant in the interstitial), the variational choice of E_l is to set it to the band energy. This, of course, gives the APW basis set enough variational freedom as the wavefunction deviates from the atomic reference. The muffin-tin approximation, however, introduces significant shifts in the calculated energy bands for all but the simplest of systems; this naturally motives one to use a *general* potential. Unfortunately, this proves to be rather difficult; different bands have different orbital characters, and by extension, experience an effective potential that differs from the spherical average used to determine the radial function $u_l(r, E_l)$. This means that setting E_l to the band energy is no longer the optimum choice and introduces the considerable difficulty in the use of a general potential.

4.2.2 The LAPW Method

The augmented wave method bloomed into prominence in the 1970s: Anderson [71] showed that the corresponding energy dependent basis set could be mapped onto one that is independent of energy. This way, the original APW method is reduced to a much simpler eigenvalue problem that has the added advantage of treating the s, p, d, and f bands in a consistent way (since the remaining shape approximation to the potential inside the atomic spheres can be removed [72]). Accordingly, the secular equation becomes linear in energy and, as a result, all eigenvalues (at a given **k**-point) are found simultaneously with a single diagonalization. Therefore, with the introduction of the energy independent basis sets, the *linearized*-APW (LAPW) method proves to be a flexible and accurate band structure method.

The basic idea is to replace the exact solutions of the radial Schrodinger equation inside the muffin-tin spheres by approximate ones at an arbitrary, but fixed energy,

$$u_l(\epsilon, r) = u_l(E_l, r) + (\epsilon - E_l)\dot{u}_l(\epsilon, r) + O((\epsilon - E_l)^2). \tag{4.35}$$

The energy derivative $\dot{u}_l(r)Y_{lm}(\hat{r})$ solves the equation

$$\left[\frac{-d^2}{dr^2} + \frac{l(l+1)}{r^2} + V(r) - E_l\right] r\dot{u}_l(r) = r u_l(r). \tag{4.36}$$

With this extension, the basis set becomes

$$\varphi(\mathbf{r}) = \begin{cases} \Omega^{-1/2} \sum_{\mathbf{G}} c_{\mathbf{G}} e^{i(\mathbf{G}+\mathbf{k})\cdot\mathbf{r}} & \mathbf{r} \in I \\ \\ \sum_{lm} [A_{lm} u_l(r) + B_{lm} \dot{u}_l(r)] Y_{lm}(\mathbf{r}) & \mathbf{r} \in S \end{cases} \tag{4.37}$$

where the B_{lm} are coefficients for the energy derivative, analogous to A_{lm}, that are chosen to ensure continuity in first derivative onto the planewave.

The energy-independent APWs are not, individually, solutions at any energy in the spheres, but they form a complete set of functions in the neighbourhood of the arbitrary energy E_l [71]. In that is the verifiable vigor of the the LAPW method; if E_l differs from the band energy ϵ inside the spheres, the linear combination (4.35) will reproduce the radial function at the proper band energy (or not far from it). In other words, the LAPW method has more variational freedom and sufficient flexibility to form a good basis set over a large energy region (about E_l) so that all eigenstates can be treated with a single (set of) E_l.

The optimum value of E_l is not known *a priori*. If it is set to the band energy ϵ, then the LAPW method would, of course, reduce to the APW method. This would, however, miss the point of making the linearization (because it would lead to a non-linear eigenvalue problem that is computationally demanding and inefficient). Instead, E_l is set to the center of the bands of *interest* (which leads to a general eigenvalue problem). For instance, suppose that one is interested in describing an eigenstate that predominantly has p-character (l=1) for a given atom. Since the error in the radial wavefunction, (4.35), is $O((\epsilon - E_l)^2$, it is beneficial to choose an E_l to keep $O(\epsilon - E_l)$ a minimum. This way, Equation (4.35) can be safely truncated after the linear term,

$$u_l(\epsilon, r) = u_l(E_l, r) + (\epsilon - E_l)\dot{u}_l(\epsilon, r).$$

Accordingly, it would seem that one needs to set E_l to the center of the p-band, and by extension, to the center of every physically important l (s-,d-, f-) band of every atom of the system; see Figure (4.5). Conditioned by this argument, one does not simply choose *one* global energy parameter, but instead, a *set* of well-chosen energy parameters.

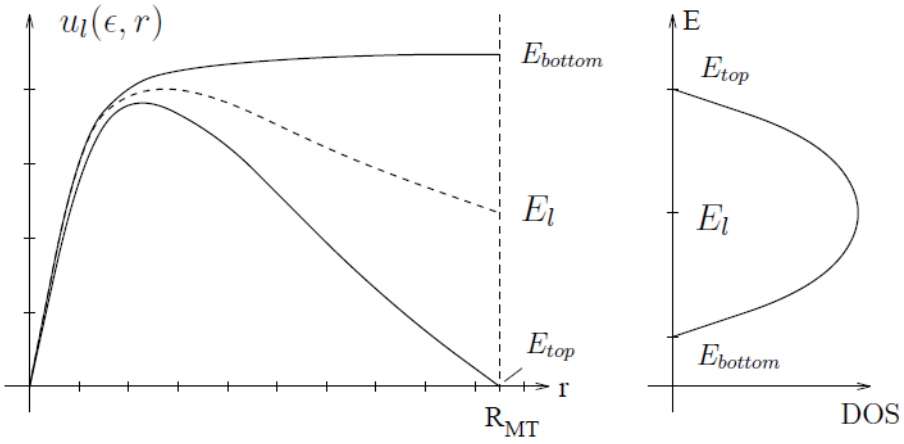


Figure 4.5: A schematic depiction of how the center of a band can be found (in WIEN2k) [73]. The value of $u_l(\epsilon, r)$ is zero in value at the bottom of the band, and has a zero slope at the top. Therefore, given a starting energy $E_l(= 0.3Ry)$, one can search up and down in small increments of energy to find the top ($u_l(\epsilon, r) = 0$) and bottom ($\dot{u}_l(\epsilon, r)$=0) of the band. If both are found, then E_l is set to the corresponding arithmetic mean.

While selecting a set of E_l is quite adequate and works well in many cases, it, unfortunately, fails miserably for many others. The reason is simple: the existence of *semi-core* states in many elements. Recall that space is partitioned, about each atomic site, into two distinct regions in the LAPW method. By definition, core states are entirely contained in the muffin tin spheres, while valence states are not; they have a significant amount of charge outside the boundary. It regularly happens that states with the same l but a different principal quantum number, n, are intermediate between core and valence states - they are semi-core states. In such a case, it is no longer clear how to choose E_l.

Consider, for instance, the example of bcc-Fe in which there is a semi-core 3p-state, and a valence state with a non-negligible amount of 4p-character. Considering both bands, it is not clear how to choose $E_{l=1}$: close to 4p, close to 3p? On the one hand, if $E_{l=1}$ is set to the semi-core (3p) energy ϵ_1, then the semi-core states are well-reproduced. But, with this choice, the errors $O(\epsilon - E_l)^2$ in the valence states would be significant (and would lie above their true positions). Alternatively, one might set the energy parameter to the desired position in the valence (4p) states. But, this too, is not satisfactory. It is often the case that this leads to spurious eigenvalues, called *ghost bands*. In short, no single choice of $E_{l=1}$ is adequate for the corresponding bands; radial wavefunctions (with E_l) are not well-suited to represent *semi*-core states.

4.2.3 The LAPW Method With Local Orbitals

It is clear that the linearization of the energy dependence of the radial wavefunction $u_l(\epsilon, r)$ does not alone provide enough flexibility to find solutions in a broader region around the chosen energy parameter E_l. The ideal solution to the dilemma of semi-core states is to add an "extension" to the LAPW basis set and use *local orbitals* [74][75][76].

The basic idea is to expand the LAPW basis with specially constructed local orbitals so that both the semi-core and valence states can be treated accurately. A local orbital is defined as:

$$\chi_{LO}(\mathbf{r}) = \begin{cases} 0 & \mathbf{r} \in I \\ \\ R_{lm}(\mathbf{r})Y_{lm}(\mathbf{r}) & \mathbf{r} \in MT \end{cases} \tag{4.38}$$

where

$$R_{lm}(\mathbf{r}) = A_{lm}u_l(r, E_{1,l}) + B_{lm}\dot{u}_l(r, E_{1,l}) + C_{lm}u_l(r, E_{2,l}). \tag{4.39}$$

Here, the coefficients A_{lm}, B_{lm}, and C_{lm} are connected by the fact that χ_{LO} should be normalized, and should vanish at the MT-boundary in value and slope. This new basis set, therefore, consists of an energy independent (linearized) APW, and a local orbital extension, that is added to the usual $u_l(r, E_l)$ and $\dot{u}_l(r, E_l)$ for certain physically important l-quantum numbers ($l \leq 3$).

Take, for example, bcc-Fe of the previous Section. In the muffin tin sphere (of a given atom), the regular LAPW radial wavefunctions, $u_l(r, E_l)$ and $\dot{u}_l(r, E_l)$, are used in conjunction with a suitable $E_{1,l}$ value that corresponds to the *highest* of the two valences states - 4p. The lower (semi-core) state, 3p, is sharply peaked at an energy $E_{2,l}$ which - given the additional flexibility of local orbitals - can now be described by a single radial wavefunction at that energy, $u_l(r, E_{2,l})$.

4.3 WIEN2k

The ideal crystal consists of a well-ordered periodic arrangement of atoms that form a solid. The translational symmetry and *cyclic* boundary conditions make it possible to treat the system as a single unit cell, which when repeated in real space generates the crystal. The unit cell may be monatomic, or it may consist of a several hundred of atoms, each of which has a nucleus and a particular number of electrons. The former problem can easily be solved on a single laptop computer, while the latter leads to large-scale grid computations. For all that, a quantum mechanical treatment is required to study the electronic structure of the many-body system. Accordingly, it can only be solved by exploiting the translational symmetry, and as such, causes the electronic wavefunctions to be of Bloch-type, where each wavefunction can be labeled by a $\mathbf{k}$ vector in reciprocal space. It follows that the periodicity in real space is defined by the $\mathbf{k}$ vector in reciprocal space, whose unit cell is the Brillouin zone.

Each electron moves in the potential (or field) of the nuclei and all the other electrons. Density functional theory provides the quantum mechanical treatment of this picture, according to which a series of one-electron Schrodinger-like equations,

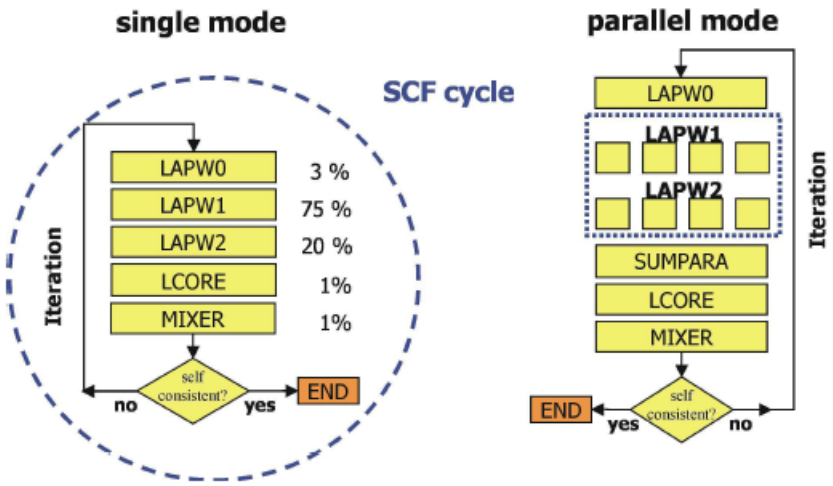


Figure 4.6: [77]

$$\left[\frac{-1}{2}\nabla_i^2 + V_i\right]\psi_i = \epsilon_i\psi_i, \tag{4.40}$$

represent the interacting electrons and nuclei. The effective potential V_i, in which a *single* electron moves, is derived from both the classical electrostatic and quantum mechanical interactions, the latter of which corresponds to the exchange and correlation effects (and is a functional of the electron density). Equation (4.40) is an eigenvalue problem where ψ_i are the eigenfunctions and ϵ_i are the eigenvalues. This problem can only be solved iteratively because:- the potential requires knowledge of the density, but the density is calculated from the sum $\psi_i^*\psi_i$ over all occupied states, which requires the effective potential to obtain the corresponding one-electron orbitals V_i. This convoluted scheme is known as the self-consistent-field (SCF) cycle and continues until some predetermined convergence criterion is achieved; see Figure (4.6).

A trial wavefunction for orbital ψ_i provides an approximate DFT solution. It is expanded in a basis set, which presently consists of augmented planewaves. Accordingly, the linearized augmented planewave (LAPW) method is the basis of the WIEN2k software package [73]. The energy expectation value calculated with any trial wavefunction, ψ_i, is an upper bound; the lower one gets in energy, the closer one becomes to the exact solution. According to the variational principle (See Appendix B), the best "solution"

for a given trial wavefunction is found by varying the coefficients of its basis set until an adequately low energy expectation value is obtained. *Minimizing* the energy this way leads to a set of linear equations. In matrix notation, this is the general eigenvalue problem $HC = SCE$ where H is the Hamiltonian, S the overlap matrix (which is a unit matrix if the basis set is orthonormal), and C the matrix containing the eigenvectors.

The SCF cycle begins with some starting electron density, generated from atomic calculations. The program LAPW0 then computes the total potential of the crystal as the sum of the Coulomb potential V_C [78], and the exchange-correlation potential V_{XC} (which is calculated numerically on a grid):

$$V = V_C + V_{XC}. \tag{4.41}$$

With this, the Hamiltonian and overlap matrices can be set up. The major effort (See Figure (4.6)) in a DFT calculation comes from solving the corresponding eigenvalue problem; there are as many different eigenvalue problems to be solved as there are **k**-points (in the IBZ). LAPW1 generate the eigenvalues and eigenvectors [79]. A synchronization step, LAPW2, is needed after the diagonalization is completed for all **k**-points. This module computes the Fermi energy E_k (which separates the occupied from unoccupied states) so that a new valence density can be found according to

$$\rho_{val} = \sum_{E_k<E_F} \psi_k^* \psi_k. \tag{4.42}$$

Added to this valence density is the core electron density ρ_{core} (calculated in LCORE). The corresponding (output) density is usually mixed with the old (input) density to generate a "new" density for the next SCF iteration [80][81]. The SCF convergence is tested according to different criteria (such as changes in total energy or charge density between the last iterations). If the convergence is better than the given threshold, the SCF cycle is stopped. Otherwise, the next cycle is started. Figure (4.7) nicely summarizes the SCF cycle in WIEN2K.

Increased computing power and disk space combined with decreasing costs have led to a significant improvement in simulating large systems on standard hardware. Today, it is reasonable to say that a single laptop PC can be routinely used in calculations that involve tens of atoms. Calculations that involve hundreds of atoms (which required powerful workstations and supercomputers about three decades ago) can be performed with a cluster

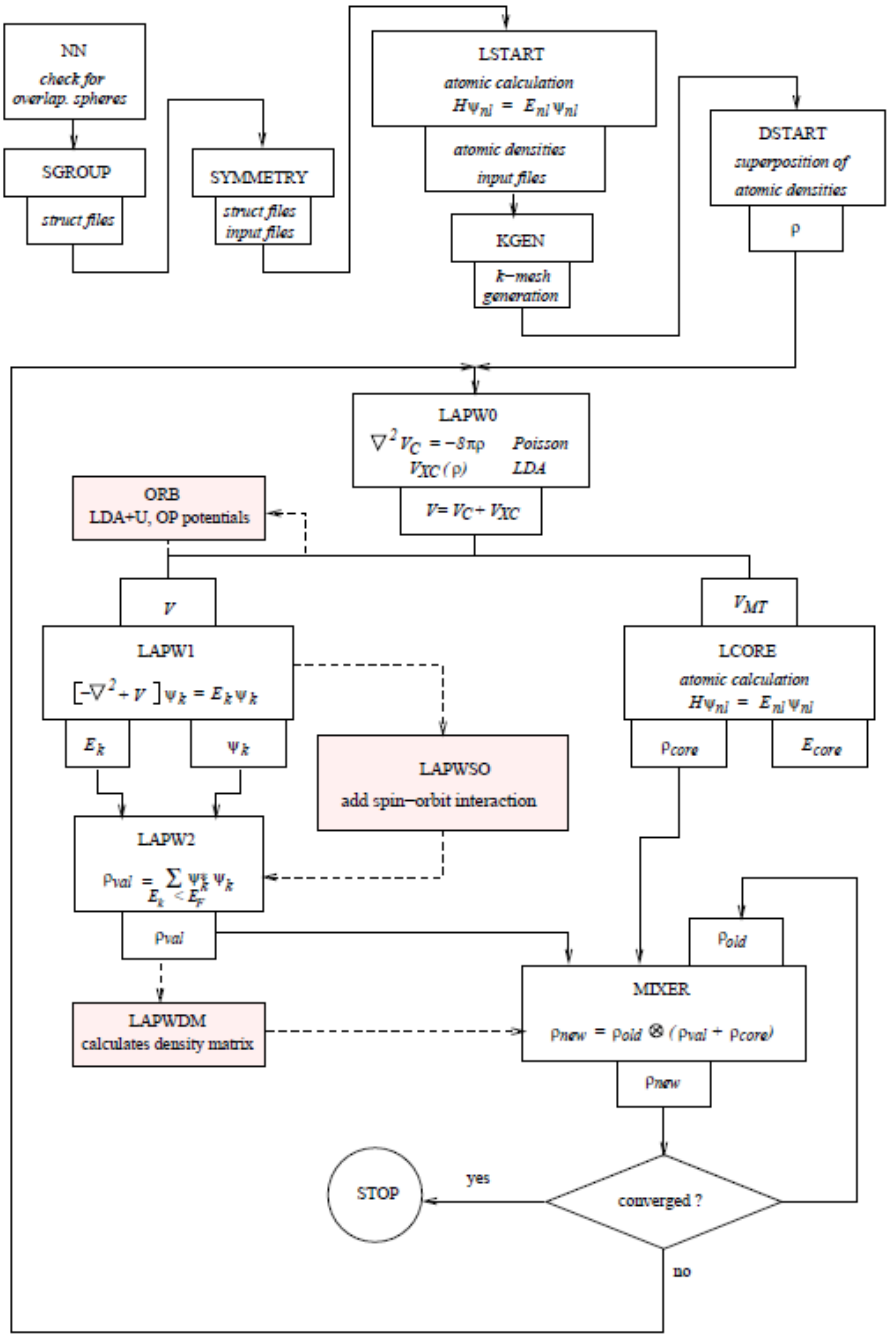


Figure 4.7: The program workflow in WIEN2k [73].

of PCs. WIEN2k takes full advantage of the available computing power by executing the most numerically intensive parts (LAPW1, and LAPW2) of the code in parallel (see Figure(4.6)). The program can run in a *coarse* grain version, where each k-point is appointed to - and computed on- a single processor, or a *fine* grain scheme (if the memory requirement is larger than that available on a single computer) [73].

The program package WIEN2k allows one to perform electronic structure calculations of crystalline solids with density functional theory (DFT). It is based on the full-potential (linearized) augmented plane-wave ((L)APW) + local orbitals (lo) method, among one of the most accurate schemes for electronic structure calculations [82]. It is for this reason that WIEN2k was chosen for this work.

Chapter 5

A Detailed Example: Electronic Properties of Si

In this chapter, the execution of a DFT calculation for a structure of simple complexity is analyzed. A clear understanding of the technical details of performing such calculations spark reliable conclusions and can assist in more complicated pursuits. Therefore the simple, yet, interesting case of silicon is lugged into analysis.

5.1 Crystallography

The properties of crystals are fairly familiar among the scientific community and have been the subject of study for hundreds of years. The peculiar science of crystallography, however, is very exact and rather indubitable. It is certainly true that to take full advantage of the theory of crystallography, the mathematics of group theory is required [83]. However, in most applications it is the features of singular cases that matter. Therefore, a full analysis of crystallography is usually inessential (and would go beyond the scope of this work). It is on this perception that its *language*, as applied to crystals, will be considered.

The objective of crystallography is concerned with the classification and determination of the crystalline structure of solids. On this wave, it is sensible to carefully ask: What is a crystal? In idealized theory, a crystal consists of atoms arranged in highly ordered patterns that periodically repeat and extend in all three dimensions [12]. This pattern can contain a single atom

(or molecule) or a group of atoms (or molecules). The important feature is the periodicity of the structure.

An abstract representation of the interior of the physical crystal is given by the translational set of mathematical points,

$$\mathbf{R} = n_1\mathbf{a} + n_2\mathbf{b} + n_3\mathbf{c}, \tag{5.1}$$

where n_1, n_2, n_3 are integers and $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{c}$ are vectors that define an appropriate coordinate system from which an arbitrary *unit* cell can be defined. The unit cell is simply a motif that which upon integral translations of its edges ($\mathbf{a}$, $\mathbf{b}$, and $\mathbf{c}$) generates an infinite array of points having identical surroundings. The corresponding set of points is called a lattice and, as previously stated, is only an abstract representation of the crystal. The process of establishing the *physical* crystal structure amounts to locating the atoms, called a basis, within one unit cell.

The seemingly easy task of defining a unit cell is a matter of considerable difficulty since there is no unique way to choose one. Any parallelepiped whose volume exactly fills space by integral translation is a valid unit cell. In the most *primitive* case, the unit cell includes within its volume precisely one lattice point located at $\mathbf{R}$. However, this choice may have the disadvantage of not displaying the full symmetry of the lattice [84]. In this respect, additional points are added and may be located at the center (I), pair of opposite faces (C), or on each face (F) of the unit cell. It therefore proves to be the case that the principle criterion for selecting a favourable unit cell is based on using symmetry.

It will become clear that in the interest of theoretical calculations, it is important to work with unit cells that complies to particularly simple rules: The cell should be the smallest, the simplest, and have the highest possible symmetry. The ornate details of how symmetry can be used to select a favourable unit cell and reduce the work required for a DFT calculation will be discussed in the next sections.

5.1.1 Symmetry

Any object possesses symmetry if some movement or *operation* on the object leaves it in an indistinguishable position. If the object is finite, it transpires that all symmetry operations can be perceived as proper (n) and improper ($\bar{n}$) rotations about certain axes of the object [84]. The second mentioned

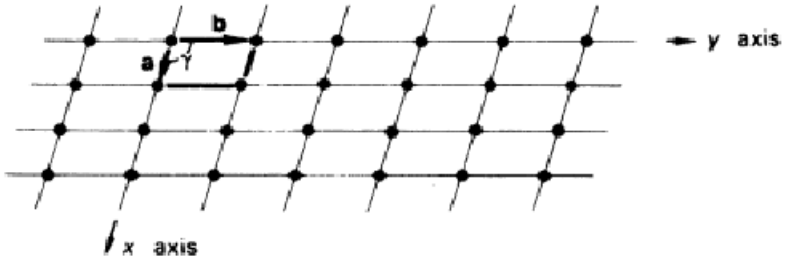


Figure 5.1: Two-dimensional section of the ab-plane of a lattice characterized by a twofold rotation symmetry about each lattice point.

of two (also known as rotoinversions) arises from a *combination* of rotation and inversion at a point. If this combination (and other stringent symmetry operations) produces a transformation that leaves at least one point of the object fixed while moving other features into itself, then the corresponding set of operations is called a *point* group [83].

The concept of a point group is applicable to finite objects and is appropriate for the description of the *shape* (or unit cell) of the crystal. Every crystal has as its geometrical basis a lattice, and since it exhibits some type of symmetry, the crystallographic convention in choosing a unit cell is to choose one whose primitive vectors (**a**,**b**, and **c**) are parallel to, or coincident with, important symmetry directions in the lattice. For instance, consider the example of a two-dimensional lattice characterized by a twofold rotation symmetry about each lattice point, as shown in Figure (5.1). A sufficient representative is an oblique parallelepiped (**a**, **b**) whose integral translation builds up the infinite lattice points. However, this choice of cell, in isolation, has the disadvantage of not displaying the full symmetry *lines* of the lattice. The higher symmetry is apparent if a larger unit cell is selected with an additional lattice point at its center, as shown in Figure (5.2). The centered cell is, therefore, preferred because the primitive vectors coincide with the point group symmetry at each point of the lattice.

Depending on the shape and symmetry of the underlying unit cell, crys-

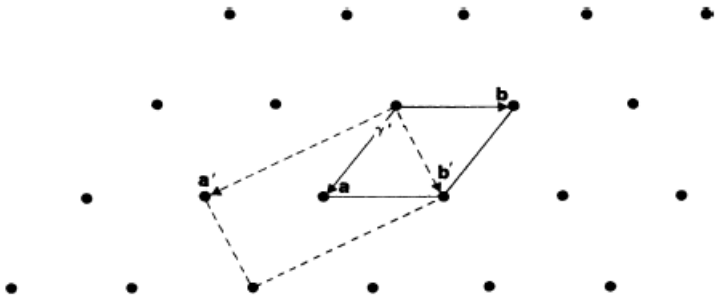


Figure 5.2: The primitive (**a**, **b**) and centered (**a**′, **b**′) unit cells of the lattice. The primitive cell has a smaller volume and less symmetry.

tals are systematically classified into one of seven hierarchical crystal systems: triclinic, monoclinic, orthorhombic, tetragonal, trigonal, hexagonal, and cubic. Considering the full symmetry and all unique possible lattice centering (P, I, C, F), there are only fourteen distinct types of unit cells, known as Bravais lattices [83]. However, a complete classification of crystals requires the study of all *translations* parallel to the three non-coplanar directions **a**, **b**, and **c**. The ensued repeat mechanism combined with the pattern motif produces indistinguishable atom-like arrangements. Such a group - including both point symmetry elements and translations of a crystal - is called a *space* group [83].

Therefore, a crystal is classified in accordance to symmetry. It may belong to one of 230 space groups, 32 point groups, 14 Bravais lattice, and 7 crystal systems. A crystal is diagrammatically represented by a delineating orderly stacked unit cell that determines which of the seven crystal systems it belongs to. Unit cells of the same shape may have additional points on their faces or at their centres. These additional points divide the 7 crystal systems into 14 Bravais lattices - which are subdivided into 32 point groups. Each point group corresponds to one of the possible symmetry operations: rotation, improper rotation, inversions, and reflections ($\bar{n} = 2$). To come full circle after considerable effort, the inclusion of translational symmetry produces the 230 space groups.

These seemingly deceptive ideas, when applied to crystals, have significant and advantageous benefits. Critical to the purposes of understanding the physical properties of materials is the ability to construct a *unique* and *efficient* representation of the lattice, and is the most basic input for a DFT calculation.

5.1.2 Crystal Structure

The structure of crystalline material is determined by the study of monochromatic radiation incident towards a portion of the crystal which upon impact results in a diffracted effect in various directions. Depending on the relative length scales involved in the experiment - noting the similarity or dissimilarity of the wavelength to the individual dimension of the target - different approaches are necessary. The incident radiation may be thought of as beams of particles (such as neutrons or electrons) or in the crystallographically dominant case, x-rays. The pertinent diffraction qualities of the radiation sources on a distant screen are their *geometrical* and *intensity* aspects.

The geometrical aspect, on the one hand, concerns the position of the diffracted beams on an intercepted pattern and only depends on the direct lattice of the crystal through the Bragg relation,

$$n\lambda = 2d \sin\theta, \tag{5.2}$$

where $n\lambda$ is an integral number wavelength and d is the separation distance between a family of the diffracted lattice planes. It is also asserted, here, that the energy of the diffracted beam be that of the incident beam so that there occurs specular reflection, defined by the Bragg angle θ. Only the size and shape of the unit cell can be determined from the intercepted geometrical diffraction pattern. The intensity of the diffracted beam, on the other hand, depends on the extent to which it interferes and is related to the atomic positions in the unit cell. The methods by which the intensity data leads to the determination of atomic positions is currently an active field of research.

Crystal Structure Determination: A Modest Overview

If the incident radiation source is thought of as x-rays, then the corresponding mathematical treatment asserts a Fourier expansion of the electron density distribution $\rho(xyz)$. This is because x-rays act in response to the local electron density in the crystal whose maxima serves to the express the centers of

atoms. The periodicity of the atoms, therefore, allows for a Fourier treatment of the electron density,

$$\rho(xyz) = \frac{1}{V}\sum_h\sum_k\sum_l F(hkl)\exp[-2\pi i(hx+ky+lz)] \tag{5.3}$$

where V is the volume of the unit cell, h, k, and l, are integral components over which the series is summed (describing the orientation of crystal planes separated by a distance d), and $F(hkl)$ is the structure factor - a to be determined coefficient.

Knowing the electron density at every point is, for all practical purposes, equivalent to knowing the crystal structure. Therefore, if the Fourier coefficients - or structure factors - were directly available from diffraction experiments, the electron density could be calculated from equation (5.3) and the crystal would be established. However, diffraction experiments only contributes to the problem the *absolute square* of the structure factor, $|F(hkl)|^2$, through the measured intensity of the reflected radiation since

$$I(hkl) \propto |F(hkl)|^2. \tag{5.4}$$

Unfortunately, this measurement is *incomplete* and the readily obtainable sets of $|F(hkl)|^2$ do not lead to crystal structure determination. This is because the structure factor is a complex number,

$$F(hkl) = |F(hkl)|\exp(i\phi) \tag{5.5}$$

and only the magnitudes are obtained from equation (5.4). In other words, the associated phase information cannot be determined in an ordinary diffraction experiment - it is systematically lost in measurement. This constitutes the greatest impasse for x-ray crystallography and a variety of techniques are available that furnish a way to success.

Performing the Fourier transforming of the electron density distribution, $\rho(xyz)$, provides a discerning and encouraging first approach to the *phase problem*: a means to calculate phase information, given a reasonably accurate initial guess of how the atoms pack in the crystal lattice. In an express manner, the structure factor can be written as:

$$F(hkl) = \sum_j f_j\exp[2\pi i(hx_j+ky_j+lz_j)] \tag{5.6}$$

where f_j is the atomic scattering factor,

$$f_j = \int\int\int \rho(x'y'z')\exp[2\pi i(hx' + ky' + lz')]dx'dy'dz', \qquad (5.7)$$

a measure of the elastic scattering amplitude of the jth atom in the unit cell. Here, the electron distribution $\rho(x'y'z')$ - whose coordinates indicate the center of the particular jth atom - are known to a good approximation by quantum mechanical methods, and therefore, the scattering factor can be numerically calculated and are veritably found in crystallographic tables.

The implication of equation (5.6) is important. If a good *guess* places the atoms in about the correct locations in the crystal, then the phases can be calculated and will be nearly correct. A useful electron density map can, thereupon, be computed by combining the observed Fourier magnitudes (5.4) with the calculated phases. Although this procedure is fast when there are only a few atoms in a unit cell, it may take months or even years for complex structures. To that end, it should be rather obvious that phase value information may be provided by a variety of techniques. For example, the Patterson function method [85], molecular replacement [86], isomorphous replacement [87], and - in my personal opinion - a particularly intriguing crystallographic analysis making use of a scanning electron microscope [88].

One might be tempted to multiply both sides of equation (5.6) by their complex conjugates, giving

$$\begin{aligned} |F(hkl)|^2 = &\Big(\sum_j f_j[\cos 2\pi(hx_j + ky_j + lz_j)]\Big)^2 \\ &+\Big(\sum_j [f_j \sin 2\pi(hx_j + ky_j + lz_j)]\Big)^2 \end{aligned} \qquad (5.8)$$

where the only unknown quantities are the positions x_j,y_j,z_j of the atoms since $|F(hkl)|^2$ are readily obtained from experiments (5.4). The truth of the situation is, however, that equation (5.8) is a set of simultaneously intractable equations and is left as an open ended problem.

5.2 Choosing a Functional

First-principle calculations, in which the only information obtained from experiment is the crystal structure, are becoming increasingly popular. Simple

models of electronic behavior (from which useful empirical models have been constructed to predict other related properties) are not irrelevant in essence but are becoming less so (at least to some degree). Today, systems of interest are large, complex, and cumbersome which limits the value of simple models, and so the electronic structure is primarily solved *ab initio*.

The significant challenge to a direct solution of the Schrodinger equation for the electrons in a crystal is exhausting because of the Coulomb repulsion between them. In DFT, this intractable problem is avoided by solving a system of non-interacting electrons but defined to have the same one-electron density as the actual system. In principle, this approach is fitted to produce the exact ground-state energy and density, but in practice, one must approximate a small but essential contribution - the exchange-correlation energy. Unfortunately, there is presently no systematic approach to approximate this (universal) functional, and so much of modern DFT research is devoted to doing just that.

Hundreds of different functionals have been suggested [89], and in any practical case, the *choice* of the functional depends on the system. It is true that a universal functional exists out of necessity (and should be attainable within the DFT framework) [57], but as of now, the quality of the calculation depends on how well this bit is approximated. Accordingly, each user must approach the title question very carefully. In this section, the key points that help with choosing a functional will be explored.

Jacob's Ladder

Before adopting a functional in practice, it is essential that the user understands the similarities and differences between the many different functionals that are in common use today. John Perdew and co-workers have described a useful classification of functionals [90] making reference to the Biblical account of Jacob, where "... he dreamed that there was a ladder set up on the earth, the top of it reaching to heaven; and the angels of God were ascending and descending on it."

Each rung of Jacob's ladder represents an approximation for the exchange-correlation energy as a functional of the electron density; higher rungs represent functionals that include more and more physical information (and typically adds accuracy to a calculation). To complete the ladder is to step to the universal functional in which the Schrodinger equation is solved exactly, as shown in Figure (5.3).

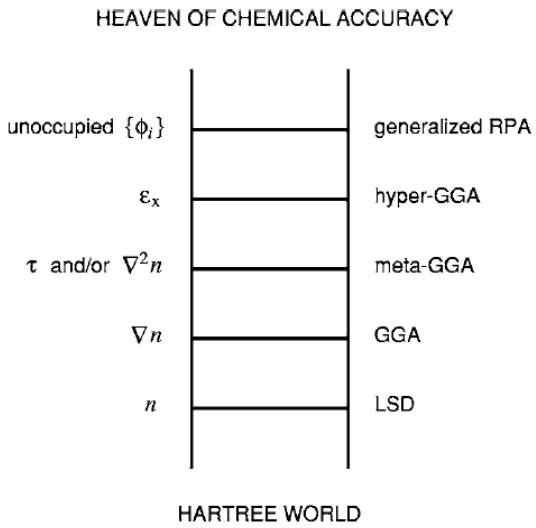


Figure 5.3: Illustration of Jacob's ladder of density functional approximations to the exchange-correlation energy. The physical constituents included in the funtionals from each rung are displayed on the left.

Functionals range from very simple to very complex. The lowest rung (and the simplest approximation) is the local density approximation (LDA). In the LDA, the exchange-correlation energy density is set at each point to that of a uniform electron gas of that density:

$$V_{XC}^{LDA} = V_{XC}^{\text{electron gas}}[n(\mathbf{r})]. \tag{5.9}$$

Different functionals of this kind were produced when this approach was initially developed, but they are practically equivalent because they all carefully and accurately describe $V_{xc}^{\text{electron gas}}(n)$. The LDA, however, does not generally get the features of the experimental band structure right. The band gap in semiconductors and insulators is commonly found to be underestimated when compared with experiment [91]; the electron density of atoms and molecules do not, in general, smoothly vary.

On the next rung are the generalized gradient approximations (GGAs). These include information on the *spatial* variation in the electron density:

$$V_{XC}^{GGA} = V_{XC}[n(\mathbf{r}), \nabla n(\mathbf{r})]. \tag{5.10}$$

With this combined information (the density and its gradient), the user can expect, with some exceptions, a functional that is typically more accurate with greater flexibility to describe real materials. Popular GGAs include PW91 [92] and PBE [93].

The third rung of the ladder is the meta-GGAs, which additionally depend on the Laplacian of the electron density, $\nabla^2 n(\mathbf{r})$. In practice, the Kohn-Sham kinetic energy density,

$$\tau(\mathbf{r}) = \frac{1}{2}\sum_{\text{occ.}} |\nabla\varphi_i(\mathbf{r})|^2 \tag{5.11}$$

includes the same physical information as the Laplacian (with a number of advantages), and so $\tau(\mathbf{r})$ may be used in the meta-GGAs instead of $\nabla^2 n(\mathbf{r})$. Examples include TPSS [94].

The next rung is the hyper-GGAs (or hybrid functionals) and is particularly important and interesting because it mixes some *exact* exchange with a GGA [95]. An essential feature of the exact exchange, which can be derived from the exchange energy density and written in terms of the Kohn-Sham orbitals,

$$E^{\text{exchange}}(\mathbf{r}) = -\frac{1}{2n(\mathbf{r})}\sum_{occ.}\int d^3r' \frac{|\varphi_i^*(\mathbf{r}')\varphi_i(\mathbf{r})|^2}{|\mathbf{r}-\mathbf{r}'|}, \tag{5.12}$$

is that it is *non-local*. This adds challenges to a calculation because when a functional is based on such a quantity (the exact exchange energy), it cannot be evaluated at one point in space without knowing the electron density at *all* points in space. In other words, it depends not only on the electron density but also on the density matrix.

The most popular of this kind is the B3LYP functional [96]. By mixing in only a fraction of exact exchange (about 25%), one can construct a very accurate functional, capable of simulating the effects of static correlation. It can be written as:

$$\begin{aligned} V_{XC}^{B3LYP} = V_{XC}^{LDA} &+ \alpha_1(E^{\text{exc.}} - V_X^{LDA}) \\ &+ \alpha_2(V_X^{GGA} - V_X^{LDA}) + \alpha_3(V_C^{GGA} - V_C^{LDA}), \end{aligned} \tag{5.13}$$

where V_X^{GGA} is the Becke 88 exchange correlation functional [97], V_C^{GGA} is the Lee-Yang-Parr correlation functional [98], and α_1, α_2, and α_3 are optimizing

parameters. The B3LYP functional has been successful in predicting the properties of small molecules, but fails to predict those of bulk materials.

Although Jacob's ladder is serviceable in classifying and characterizing functionals, there are other important, relevant, and specific considerations that need to be considered when making distinctions between the many different functionals that are in frequent use today. Possibly the most significant of these is whether a given function is non-empirical, a little empirical, or over-empirical. At each rung of the ladder, each of these flavors of functionals has been developed.

Non-empirical functionals have been constructed to satisfy only general rules of quantum mechanics and special limiting conditions. They are not fit to any molecule or crystal and try to satisfy as many exact conditions as possible. Successful functionals of this type tend to have errors that are pretty systematic. The LDA, for instance, always overbinds: the lattice parameters are underpredicted, while the cohesive energy is overpredicted. There is still a bias toward non-empirical functionals; any derived functional outside this flavor is ignored if it is inaccurate.

Empirical functionals, in contrast, are fitted by experts to selected *ab initio* data. They contain some parameters that have been introduced and adjusted during the fitting process. This can often facilitate functional construction and typically reduce errors on systems similar to those which it was designed for. However, these errors, while smaller in magnitude, are typically unsystematic. They are, therefore, less reliable when applied under new conditions.

5.2.1 The Band Gap Problem

The most fundamental property of a periodic solid is its band gap. It vanishes for metals and is positive for semiconductors and insulators. However, a long-standing challenge for DFT is its inability to predict the band gap; the so-called band gap problem [99][100] .

In the Kohn-Sham method of DFT, electronic structure calculations of periodic systems are performed by solving the equations

$$\left[-\frac{\hbar^2}{2m}\nabla^2 + V_{ext}(\mathbf{r}) + V_H(\mathbf{r}) + V_{XC}(\mathbf{r})\right]\psi_i(\mathbf{r}) = \epsilon_i\psi_i(\mathbf{r}). \tag{5.14}$$

for the one-electron wavefunctions $\psi_i(\mathbf{r})$. In this equation, the first two terms

are the external and the Hartree potential, and the last is the exchange-correlation term (which, of course, must be approximated). The exchange-correlation potential is obtained as the functional derivative:

$$V_{XC}(\mathbf{r}) = \frac{\delta E_{XC}[n(\mathbf{r})]}{\delta n(\mathbf{r})} \tag{5.15}$$

where E_{XC} can be linearly decomposed as

$$E_{XC} = E_X + E_C \tag{5.16}$$

so that individual expressions for E_X and E_C can be found [101]. To solve the Equations (5.14), an explicit expression for $E_{XC}[n(\mathbf{r})]$ is needed. The most popular approximations are the LDA and the PBE functionals (which reproduce rather well the band structure of metallic systems) but systematically underestimates the band gaps of semiconductors.

Modified Becke-Johnson Potential

Much work has been devoted to improve the predictions of band gaps. As a workable empirical solution to this problem, Blaha *et al.* [102] have proposed the mBJ potential, which is a modification of the exchange-correlation potential of Becke and Johnson (BJ) [103]. It is given by

$$V_X^{mBJ}(\mathbf{r}) = cV_X^{BR}(\mathbf{r}) + (3c-2)\frac{1}{\pi}\sqrt{\frac{5}{12}}\frac{\sqrt{2t(\mathbf{r})}}{n(\mathbf{r})} \tag{5.17}$$

where

$$n(\mathbf{r}) = \sum_{i=1}^{N} |\psi_i(\mathbf{r})|^2 \tag{5.18}$$

is the electron density,

$$t(\mathbf{r}) = \frac{1}{2}\sum_{i=1}^{N} [\nabla\psi_i^*(\mathbf{r}) \cdot \nabla\psi_i(\mathbf{r})] \tag{5.19}$$

is the Kohn-Sham kinetic energy density, and V_X^{BR} is the Beck-Roussel exchange potential [104]. The value of c is expressed in such a way that the LDA exchange potential is recovered from Equation (5.17) in the high-density limit:

$$c = \alpha + \beta \left(\frac{1}{V_{\text{cell}}} \int d^3 r \frac{|\nabla n(\mathbf{r}|}{n(\mathbf{r})} \right)^{1/2}. \tag{5.20}$$

The value of c is calculated for every system under study; V_{cell} is the unit cell volume, and α and β are free parameters chosen to produce the best fit to experimental results. In the mBJ potential, their empirical values are $\alpha = -0.012$ and $\beta = 1.023$ bohr$^{1/2}$. With the mBJ potential, studies have shown that the band gaps of insulating and semiconducting systems can be described very accurately with an effort comparable to that of the LDA or GGA [105][106].

A particular feature of this potential is that it cannot be obtained as the derivative of a functional, namely $V_{XC} = \delta E_{XC}[n]/\delta n$. Since no energy functional is defined within this framework, the potential cannot be used for the calculation of forces which act on the nuclei, and therefore, no consistent optimization is possible. This shortcoming is the result of its empirical nature. As a possible solution to this problem, Tran and Blaha [102] suggested starting with a GGA or LDA optimization and introduce the obtained lattice parameters into the calculation (that of which uses the mBJ potential). The empirical potential, however, still gives predictions that compare acceptably well with experiment.

5.3 Numerical Data Exploring

The ground-state electron density of an arrangement of atoms as defined by DFT is fixed by the solution to Equation (5.14). When this problem is actually solved on a computer, a series of numerical approximations are made. On the one hand, integrals in multidimensional space are evaluated by studying the to-be-integrated function at a limited collection of points, while, on the other hand, solutions that are principally expressed as infinite sums must be truncated to finite sums; and so forth. With more computational resources, it is always possible to obtain a better and better solution (one that more closely reflects the exact ground state solution). However, with limited resources, it is vital to obtain a solution that is accurate but does not misuse computer time - the solution must be well-converged and efficient.

The concept of convergence is fundamental to any DFT calculation, and as such, one must continually question the rigor of the approximations used

to obtain the solution. In this section, the pertinent features required to perform well-converged calculations will be explored.

Numerical Integration

The discussion of the concept of numerical convergence occurs not in the familiar three-dimensional space (where atom positions are specified), but rather in reciprocal space. It happens that it is much more convenient to solve the mathematical problems posed by DFT in terms of $\mathbf{k}$ than it is in terms of $\mathbf{r}$. This is because a great deal of the work in performing DFT calculations amounts to evaluating integrals of the form,

$$\bar{g} = \frac{V_{\text{cell}}}{8\pi^3} \int_{BZ} g(\mathbf{k}) d\mathbf{k}, \tag{5.21}$$

which are, indeed, defined in reciprocal space (and integrated only over the possible values of $\mathbf{k}$ in the Brillouin zone). Instead of exploring in detail *where* integrals such as these come from (see Chapter 4), we will instead briefly explore *how* they can be numerically evaluated.

The basic problem in numerical integration is to calculate an approximate solution of a definite integral,

$$\int_a^b f(x)dx,$$

to a given degree of accuracy. A simple way to approximate this integral is to break up the interval [a,b] into pieces of, say, equal sizes - for which the function is regarded as a straight line between the endpoints - and calculate the area under the curve. This process is known as the trapezoidal method, and it can be shown that:

$$\int_a^b f(x)dx \approx \frac{\Delta x}{2}\big[f(x_0) + f(x_n) + 2\sum_{i=1}^{n-1} f(x_i)\big],$$

where [a,b] is broken up into n subintervals of width Δx.

As a measure of accuracy, the simple integral $\int_{-1}^{1}(\pi x/2)\sin(\pi x)dx$ is numerically drawn out via the trapezoidal method. This integral can be evaluated exactly "by-parts" and, therefore, becomes useful as a benchmark to compare the proposed method of evaluation. Table (5.1) [37] shows some of the corresponding results. As one might suspect, when the resolution of the

n	Trapezoidal Method	Legendre Quadrature Method
2	0.6046	1.70605
3	0.7854	0.8793
4	0.8648	1.0080
5	0.9070	0.9997

Table 5.1: Approximations to the integral $\int_{-1}^{1}(\pi x/2)\sin(\pi x)dx = 1$ using the trapezoidal method and the Legendre quadrature method.

subinterval widths decreases (that is, for larger n) the method becomes more and more accurate. If, for instance, the integral were to be evaluated to an accuracy of 1%, then the results in Table (5.1) suggests that the value of n must be much greater than 5.

The features that contribute to the minimalist design of the trapezoidal method - the n equally spaced subinterval widths and (apart from the end-points) the equally weighted evaluation of $f(x)$ - are not essential. In fact, neither of these conditions are even desirable. There exist more complex classes of integration methods [107] which significantly increases the convergence efficiency. The Gaussian quadrature methods, for instance, have the form

$$\int_{-1}^{1} f(x)dx \approx \sum_{i=1}^{n} w_i f(x_i)$$

where the integration points x_i are related to roots of orthogonal polynomials and the weights w_i are related to integrals involving these polynomials. When the domain of integration for such a rule is taken as [-1,1], the method is called Legendre quadrature. Table (5.1) shows the corresponding first few results from applying the method. In remarkable contrast to the trapezoidal method, the results converge much quicker; in this case, the error in the integral is less than 1% for n greater than 3.

The principal features of numerical integration can be summarized in three relevant points. First, integrals can be approximated by taking a kind of weighted sum of values of the (to be integrated) function at a set of discrete points. Secondly, in the limit of infinite discrete points, numerical methods of this type will converge to the exact result. And finally, the rate at which the

numerical method converges dramatically depends on the various weighting choices used in the sum.

5.3.1 Brillouin Zone Sampling

For a periodic system, Bloch's theorem indicates that any wavevector $\mathbf{k}$ of a given state, $\psi_{\mathbf{k}}$, can be confined to the first Brillouin zone - the reciprocal unit cell [12]. By extension, infinitely extended integrals in reciprocal space can be replaced by integrals over the (finite) first Brillouin zone. But, the first Brillouin zone, in principle, would still need to be sampled with an infinite density of $\mathbf{k}$-points. This is problematic because much of the computational effort in a DFT calculation involve integrating periodic functions of the wave vector over the Brillouin zone; (5.21). Accordingly, the question of how to evaluate these integrals have been carefully answered: In any given calculation, integrals of this type are replaced by weighted sums of the integrand using a discrete set of points (in the BZ) called the $\mathbf{k}$-point mesh. That is,

$$\bar{g} = \frac{V_{\text{cell}}}{8\pi^3} \int_{BZ} g(\mathbf{k}) d\mathbf{k} = \sum_j w_j g(\mathbf{k_j}), \tag{5.22}$$

where w_j are the weighting factors (and $g(\mathbf{k})$ is the Fourier transform of $g(\mathbf{r})$).

Choosing a sufficiently dense mesh of $\mathbf{k}$-points is not as straightforward as might be expected. There are two important qualities to consider when sampling reciprocal space. First, it is not the number of $\mathbf{k}$-points that is important, *per se*, but rather their spacing in reciprocal space. The set of *special* points for Brillouin Zone integrations are commonly generated using the Monkhorst-Pack method [108]. The coordinates of these points are given by

$$\mathbf{k}_j = x_{1j}\mathbf{b}_1 + x_{2j}\mathbf{b}_2 + x_{3j}\mathbf{b}_3. \tag{5.23}$$

where $\mathbf{b}_i$ are reciprocal lattice vectors that span the Brillouin zone, and

$$x_{ij} = \frac{l_i}{n_j}, i = 1, \ldots, n_j, \tag{5.24}$$

where l_i are *lengths* of reciprocal lattice vectors, and n_j is an integer that determines the number of points in the set. The corresponding $\mathbf{k}$-point mesh is homogeneously spread out over the whole Brillouin zone by translations of

the reciprocal vectors. Note that a large real-space unit cell corresponds to a *small* reciprocal-space unit cell,

$$V_{\mathrm{BZ}} = \frac{8\pi^3}{V_{\mathrm{cell}}}. \tag{5.25}$$

This means that only a small **k**-point mesh would be necessary to sample the BZ (because the number of k-points depends on the *size* of the Brillouin zone and the *spacing* of **k**-points). A small Brillouin zone (of volume V_{BZ}) therefore corresponds to a small mesh; the volume of k-space per allowed **k** value is

$$\Delta\mathbf{k} = \frac{8\pi^3}{V_{\mathrm{BZ}}}.$$

It might happen that for a very large simulation with many atoms per cell (say 40-60), a single **k**-point may be sufficiently dense to perform the calculation [73].

Secondly, the weight of the **k**-points account for the point group symmetry of the system (which map equivalent k-points to each other) and can significantly reduce the sampling density. As an artificial example, consider a single band sampled on a $1 \times 1 \times 3$ **k**-point mesh centered on the origin. In principle, the $E(\mathbf{k})$ is computed for all three **k**-points:

$$\begin{aligned} E_1 &= E(0, 0, 1/3) \\ E_2 &= E(0, 0, 0) \\ E_3 &= E(0, 0, -1/3). \end{aligned}$$

The total energy of the state is then given by

$$E = (E_1 + E_2 + E_3)/3. \tag{5.26}$$

In other words, the total energy has been approximated as a kind of weighted sum of values at a set of discrete points:

$$E = \sum_j w_j E_j = w_1 E_1 + w_2 E_2 + w_3 E_3. \tag{5.27}$$

In this case, the weight of each **k**-point is $1/3$. But, considering the symmetry of the system - time-reversal: $E(0, 0, 1/3) = E(0, 0, -1/3)$ - one can see that it would be necessary only to compute

$$E_1 = E(0, 0, 1/3)$$
$$E_2 = E(0, 0, 0),$$

with the proviso that $w_1 = 2/3$ because E_2 represents two **k**-points in reciprocal-space. Accordingly,

$$E = \frac{2E_1}{3} + \frac{E_2}{3}, \tag{5.28}$$

gives the same result as that of (5.26).

Using this weighting scheme with the symmetry of the system has reduced the number of **k**-points from three to two while maintaining an equivalent quality of calculation. These symmetries mean that integrals in reciprocal space do not need to be evaluated using the entire Brillouin zone, but rather, they can be evaluated using only a reduced portion of the zone that can then be extended without approximation to fill the entire Brillouin zone using symmetry. This reduced region is called the *irreducible Billouin zone* (IBZ). Accordingly, for very symmetrical systems, using just the IBZ greatly reduces the numerical effort in a calculation.

It should be stated that other conventions for choosing a **k**-point mesh exist. Chadi and Cohen [109], for instance, have proposed another convention for choosing a **k**-point mesh based on a recursive scheme, and has been applied to slab calculations by Cunningham [110]. Most DFT packages have various options for choosing **k**-points.

5.3.2 Size of Basis Set

Kohn and Sham [59] put forth a prescription for determining the ground state energy and density (given the exchange-correlation functional). In particular, they showed that the true density is produced by the self-consistent solution of a set of single particle Schrodinger-like equations,

$$\left[-\frac{\hbar^2}{2m}\nabla^2 + V_{ext}(\mathbf{r}) + V_H(\mathbf{r}) + V_{XC}(\mathbf{r})\right]\psi_i(\mathbf{r}) = \epsilon_i\psi_i(\mathbf{r}). \tag{5.29}$$

known as the Kohn-Sham (KS) equations. In this framework, the many-body problem is far easier to solve. Instead of having to consider all the individual

electron wave functions of the system simultaneously, one can alternatively determine the solution to a series of single particle Schrodinger equations [16]; the KS orbitals with different wave vectors, $\mathbf{k}$, are only indirectly coupled through the density dependent potential (5.14). Accordingly, the single particle equations can be solved independently on a grid of sampling points in the symmetry irreducible wedge of the Brillouin zone.

Recall that Bloch's theorem states that the solution of the Schrodinger equation for a given unit cell has the form

$$\psi_{\mathbf{k}}(\mathbf{r}) = \exp(i\mathbf{k}\cdot\mathbf{r})u_{\mathbf{k}}(\mathbf{r}), \tag{5.30}$$

where $u_{\mathbf{k}}(\mathbf{r})$ is periodic in space with the same periodicity as the lattice. Given that the lattice has a very specific periodicity, there is only a particular set of plane waves that can describe the solutions, $\psi_{\mathbf{k}}(\mathbf{r})$. They are:

$$u_{\mathbf{k}}(\mathbf{r}) = \sum_{\mathbf{G}} c_{\mathbf{G}} \exp[i\mathbf{G}\cdot\mathbf{r}] \tag{5.31}$$

where the summation is over the reciprodal vectors. Bringing the above equations together gives

$$\psi_{\mathbf{k}}(\mathbf{r}) = \sum_{\mathbf{G}} c_{\mathbf{k}+\mathbf{G}} \exp[i(\mathbf{k}+\mathbf{G})\mathbf{r}]. \tag{5.32}$$

In other words, the solutions *can* be solved by sampling the Brillouin zone.

DFT methods are classified according to the choice of the basis set in the construction of the KS orbitals,

$$\psi_i(\mathbf{r}) = \sum_{\alpha} c_{i\alpha}\phi_{\alpha}(\mathbf{r}). \tag{5.33}$$

Having chosen the basis functions $\phi_{\alpha}(\mathbf{r})$ (see Chapter 4), the problem of "solving" Equation (5.29) amounts to finding the coefficients $c_{i\alpha}$ that minimize the total energy. This can be done with standard matrix techniques as an eigenvalue problem. For a given KS orbital ψ_i, this leads to

$$(\mathbf{H} - \epsilon_i\mathbf{S})\mathbf{c}_i = 0, \tag{5.34}$$

where $\mathbf{H}$ and $\mathbf{S}$ correspond to the matrix elements of the Kohn-Sham Hamiltonian and the overlap matrix elements in the chosen basis states respectively; ϵ_i is the energy of the given orbital.

The rank of the square matrices **H** and **S** is equal to the number of basis functions n_i in Equation (5.33). Therefore, diagonalizing the Hamiltonian matrix **H** would lead to n_i eigenvalues and n_i sets of coefficients $\mathbf{c}_i$ that express each of the n_i eigenfunctions in the given basis set (5.33). The larger the basis set, the better the approximation to the KS orbital. However, with a larger n_i, the diagonalization of the Hamiltonian matrix would be more time-consuming; the **S** matrix is a unit matrix if the chosen basis set is orthornomal.

It is possible and beneficial to adopt a different basis set for each **k**-point (see Chapter 4). This will lead to a more efficient basis set and lessens the difficulty in computing matrix elements in the eigenvalue problem. There are, however, as many different eigenvalue problems to solve as there are **k**-points in the irreducible Brillouin zone, which corresponds to a significant and heavy computer problem. What is more, given a sufficiently dense **k**-point mesh, evaluating Equation (5.33) at a single point in reciprocal space involves a summation over an infinite number of possible values. This condition - infinite in principle - is not promising for practical DFT calculations.

In practice, one must work with a finite set of basis functions $\phi_\alpha(\mathbf{r})$; one that is large enough to be accurate - but only just. This introduces one more parameter that must be defined whenever a DFT calculation is performed. In Wien2k, it is the $R_{mt}K_{max}$, a quantity that has not been defined hitherto [73]. The implication is simple: one must perform some testing of the elementary system as a function of basis set size (or $R_{mt}K_{max}$).

5.3.3 Convergence Tests

It is imperative that one determines a well-balanced basis set size and a dense enough **k**-mesh before executing and interpreting large calculations. To achieve this without taking into account the numerical intricacies present in any calculation has adverse effects. Firstly, if less precision than necessary is used, then one risks producing nonsense results instead of physics. It is in this sense that numerical rigor is not an option but a priority. It is the user's responsibility to specify the number of **k**-points and basis set size set to ensure well-converged quantities of interest. Secondly, if there is a higher than necessary precision used in the calculation, then one risks wasting computer time. Admittedly, this is a softer issue but it is still important. Increasing the basis set size by 30%, for instance, increases the computing time by a factor of about two [73]. This means that a large-scale calculation will run

No. of k Points (BZ)	Total Energy (Ry)	No. of **k** Points (IBZ)
100	-1160.1294	8
200	-1160.1385	10
400	-1160.1425	20
800	-1160.1431	35
1600	-1160.1432	56
3200	-1160.1432	104
6400	-1160.1432	195
12800	-1160.1432	364

Table 5.2: The total energy results of Si as a function of k points generated using the Monkhorst-Pack method.

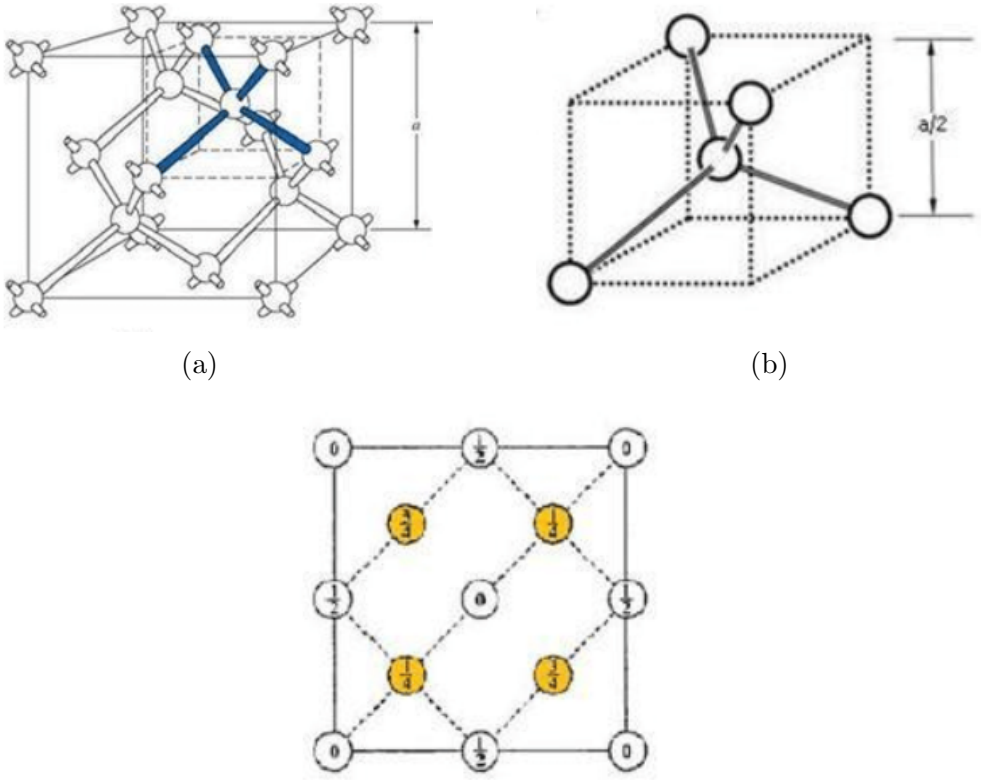
for two weeks instead of one.

Number of k-Points

From the discussion of numerical integration above, it is clear that a calculation using M **k**-points will give a more precise result than a calculation using N **k**-points if $M > N$. But, in practice, how does one choose the number of points? This is a difficult question to answer in general, and as such, only general guidelines can be given [73]: Metallic systems require a fine mesh while insulators are often okay with very few. Small unit cells require a dense mesh, while large unit cells do not (etc). The important quality ascribed to all systems under study is numerical convergence - it is the common requirement.

Table (5.2) shows the data from a series of calculations that we have performed for the semiconductor Si which crystallizes in the 'diamond structure'. The space group is Fd3m and is derived from the face centered cubic (FCC) lattice by attaching two identical atoms at coordinates $(0, 0, 0)$ and $(0.25, 0.25, 0.25)$ associated with each lattice point, as shown in Figure (5.4). Each calculation listed in Table (5.2) used **k**-points defined using the Monkhorst-Pack approach. The results from this table are shown graphically in Figure (5.5).

Consider the total energies (in Ry) listed in Table (5.2) and plotted in Fig-



(a) (b)

(c)

Figure 5.4: The crystal structure of diamond. (a) The lattice constant is given by a and the tetrahedral bond arrangement of the Si atoms is shown in blue. (b) The tetrahedral structure of the closest neighbors in the lattice. (c) Atomic positions in the cubic cell of the diamond structure projected on a cube face; fractions denote height above the base in units of a cube edge, the points at 0 and 1/2 are on the FCC lattice; those at 1/4 and 3/4 are on a similar lattice displaced along the body diagonal by 1/4 of its length [111].

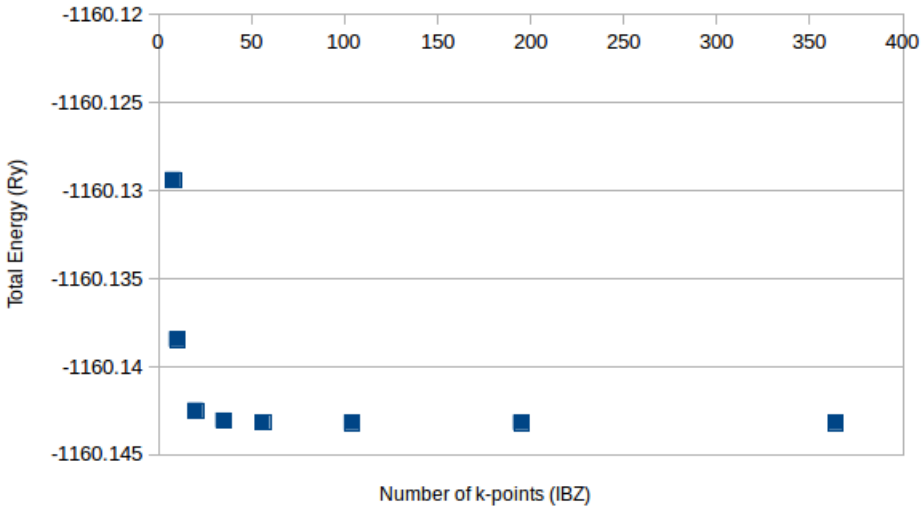


Figure 5.5: The convergence of the total energy with respect to the number of k points.

ure (5.5). Notice that when $M > 1600$, the energy appears to be independent of the number of **k**-points - an indication that the calculation is numerically well-converged. Indeed, for a smaller number of **k**-points ($M < 1600$), the energy varies with the number - an indication that the calculation is not well-converged.

The determinantal equation of maximum complexity corresponds to the maximum number of **k**-points in the full Brillouin zone. Table (5.2) shows how useful symmetry is in reducing the labor required to perform a DFT calculation. Any operation that leaves at least one point of the crystal invariant also transforms $\epsilon(\mathbf{r})$ into itself. This means that for a given symmetry element of the crystal, there will be degenerate energy levels. Instead of submitting all 100 **k**-points to a calculation, one can, for instance, use 8 **k**-points instead (as seen in Table (5.2)). The higher the symmetry of the point in the zone, the lower the degree of the equation to be solved.

Number of Basis Set Elements

The discussion above, however, has introduced another parameter that must be defined before a DFT calculation can be precisely performed: The size of the basis set. In wien2k, the size (and quality) is determined by $R_{mt}K_{max}$

- the product of the smallest atomic sphere radius (R_{mt}) times the largest **k**-vector (k_{max}) of the plane wave expansion of the wavefunction,

$$\varphi(\mathbf{r}) = \Omega^{-1/2} \sum_{\mathbf{K}} c_{\mathbf{K}} e^{i(\mathbf{K}+\mathbf{k})\cdot\mathbf{r}}.$$

Recall that space is divided into distinct regions: non-overlapping spheres about each atomic site and the remaining volume between the spheres. The dual representation of the wavefunction (Chapter 4)

$$\varphi(\mathbf{r}) = \begin{cases} \Omega^{-1/2} \sum_{\mathbf{K}} c_{\mathbf{K}} e^{i(\mathbf{K}+\mathbf{k})\cdot\mathbf{r}} & \mathbf{r} \in I \\ \\ \sum_{lm} A_{lm} u_l(r) Y_{lm}(\mathbf{r}) & \mathbf{r} \in S \end{cases}$$

is not guaranteed to exactly match over the entire spherical surface (of radius R_{mt}). This must be strictly imposed and introduces expansion coefficients A_{lm} (and B_{lm} in the LAPW method) to complete the matching. In principle, an infinite amount of coefficients are required to create a perfect matching, but in practice, this number must be truncated at some value l_{max}. But what is a reasonable choice?

For a given l_{max}, the spherical harmonics $Y_{lm}(\theta, \phi)$ can have at most $2l_{max}$ points for which it is zero (for a fixed m) along a great circle (defined by: $\theta = 0 \rightarrow 2\pi$ and radius R_{mt}) [112]. The corresponding *nodes* per unit length is given by

$$\frac{2l_{max}}{2\pi R_{mt}} = \frac{l_{max}}{\pi R_{mt}}. \tag{5.35}$$

The idea is simple enough: If a planewave should match at the boundary, then it should have a similar number of nodes per unit length. Given that the planewave with the shortest period has $2/(2\pi/k_{max}) = k_{max}/\pi$ nodes per unit length, one can write

$$\frac{l_{max}}{\pi R_{mt}} = \frac{K_{max}}{\pi}, \tag{5.36}$$

which leads to the robust condition:

$$R_{mt} K_{max} = l_{max}. \tag{5.37}$$

This allows one to determine a good l_{max} for a given K_{max}, and is a good indicator of the quality of the basis set. Clearly, larger values of $R_{mt}K_{max}$

$R_{mt}K_{max}$	Smallest Atom
3.0	H
4.5	Li
5.0	Be, B, Si
5.5	C, P
6.0	N, S
6.5	O, Cl, Na, K, Rb Cs, Mg, Ca, Sr, Ba, Al
7.0	F
7.5	Sc-Cr, Ga-Br, Y-Mo
8.0	Mn-Zn, Ru-Cd, In-I, La, Ce, Hf-Re
8.5	Os-At, Pr-Lu, Ac-Lr

Table 5.3: Approximate values for $R_{mt}K_{max}$ per element [73].

leads to more accurate results, but also increases the amount of necessary computer time (since the basis scales with $(R_{mt}K_{max})^3$).

To push the discussion just a bit further, consider the underlying physics. For a given atom, the wavefunctions become more rapidly oscillatory, in both angular and radial directions, as one approaches the nucleus. Therefore, a fixed $R_{mt}K_{max}$ inherently improves the planewave resolution (K_{max}) at the boundary (R_{mt}) as R_{mt} is *decreased*, as required to fit the more rapid oscillations in the wavefunction at R_{mt}. Conversely, as the boundary radius *decreases*, (a fixed) $R_{mt}K_{max}$ naturally decreases the planewave resolution to accommodate the smoother oscillations in the wavefunction (because the matching is not as coarse and requires fewer coefficients A_{lm} (and B_{lm})).

Are the oscillations twice as rapid at 1/2 Rmt? Well, no, of course not. But $R_{mt}K_{max}$ gives a far better indicator of numerical convergence as a single number than K_{max} does alone. Table (5.3) provides approximate guidelines for a low to medium convergence [73].

Table (5.3) is not meant as a substitute for explicit testing. It simply provides a quick guideline for choosing $R_{mt}K_{max}$ in the approximate region that should be tested. Assume that the kind of accuracy needed for the Si

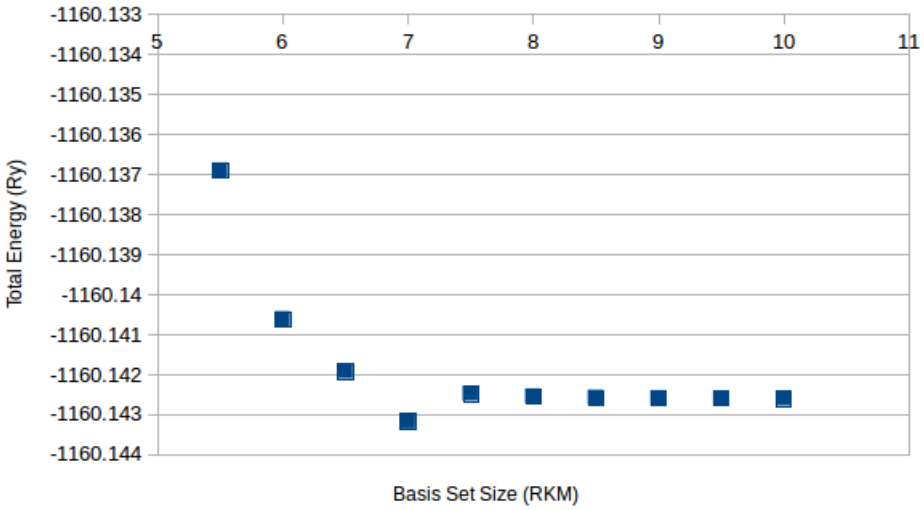


Figure 5.6: The convergence of the total energy with respect to the size of the applied basis set and fixed **k** point mesh.

calculation requires 56 k-points in the IBZ (1600 k-points in the full BZ). Accordingly, this value is fixed, and the value of $R_{mt}K_{max}$ is varied from 5.5 (a small basis set) to 10.0 (a very large basis set), in increments of 0.5. Figure (5.6) shows the total energy as a function basis set size. It is shown that for $R_{mt}K_{max} > 7.5$, the total energy is almost independent of the size of the basis set. In other words, the calculation is numerically well-converged.

The basis size and **k**-mesh density only determine the *numerical* accuracy. Given a very large basis set and a very dense mesh, one can only obtain the exact value for a particular exchange-correlation functional. It does not ensure that the exact mathematical solution will be close to the exact value of a property as it appears in nature.

5.4 Electronic Structure Calculations of Si

Figure (5.7a) shows our calculated density of states (DOS) for Si from a DFT calculation using the PBE [93] exchange-correlation functional, 680 k points (in the IBZ), and an $R_{mt}K_{max}$ value of 8.0. The DOS is divided into two separate regions: the valence band and the conduction band. The valence band is the collection of all occupied electron states, while all the states in

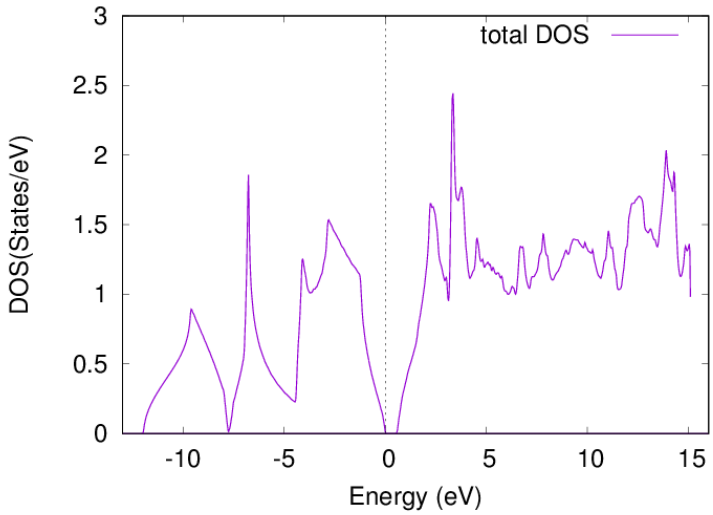
Energy Gap in eV		
Present Work		Available Data
PBE	**mJB**	Experimental
0.67	1.16	1.1

Table 5.4: The calculated band gap values for Si using the PBE exchange-correlation functional and the mBJ potential. Notice that the PBE functional underestimates the band gap by about 50%.

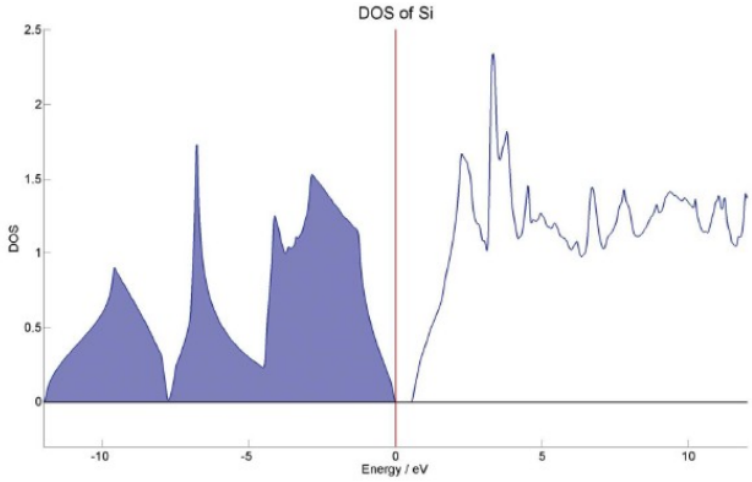
the conduction band are unoccupied. The region of energy that separates the bands contain no electronic states at all; this is the band gap. While the calculation correctly predicts the existence of the band gap, its width is not correct. The experimentally observed band gap is 1.1 eV [12], and the calculated band gap is 0.67 eV. Recall that this underestimation is a systematic error with DFT. Indeed, the well-converged results only correspond to the particular exchange-correlation functional (PBE) used in the calculation. Since the exact functional is not used (or known) for the calculation, there is no reason for the results to be close to the value of that which appears in experiment.

Table (5.4) reflects the importance of the particular functional used in the calculation. In Section 5.2.1, a workable empirical solution to the "band gap problem" was discussed. Notice that the modified Becke-Johnson (mBJ) potential correctly predicts the width of the band gap. Given this potential, we performed a new series of calculations. Figure (5.8) illustrates the corresponding DOS, which correctly predicts the band gap width for Si. Each user must approach every calculation carefully and spend time understanding the state-of-the-art literature (in their field) to choose an appropriate functional for their problem of interest.

The properties and characteristic evolution of the electronic states in reciprocal space are nicely represented in the DOS. A more subtle view of the material's electronic structure is possible by studying the band structure. Each state has an energy that changes with $\mathbf{k}$ - they form bands. The band structure describes the range of energies that an electron within the solid may or may not have along a series of lines in reciprocal space. Attention must be devoted to calculations of this type to the position of $\mathbf{k}$-points. Accordingly, the electronic states are evaluated at special high-symmetry points (in the first Brillouin zone) appropriate to the point group symmetry of solid.



(a)



(b)

Figure 5.7: Calculated electron DOS for bulk Si with the PBE exchange-correlation functional: (a) our calculation (b) in literature [113].

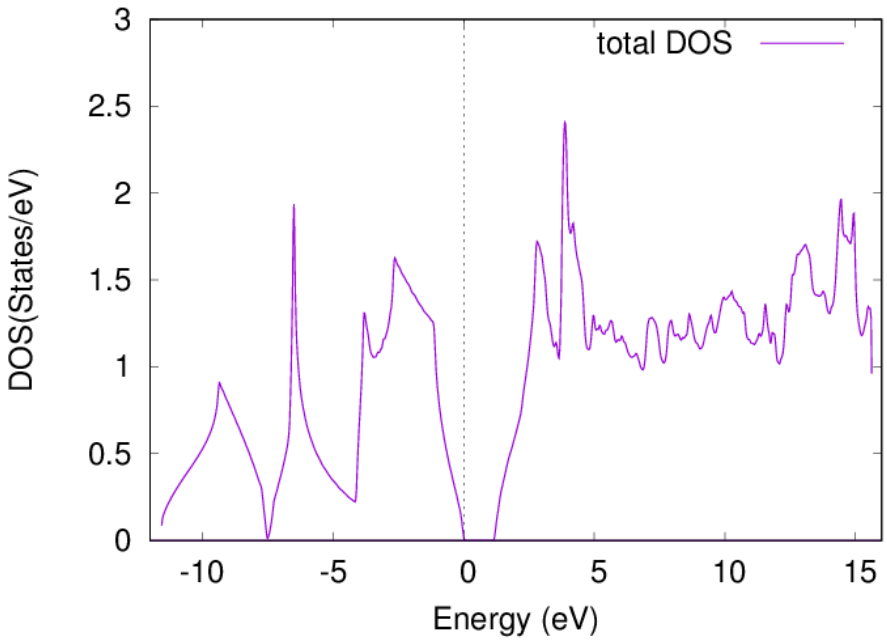


Figure 5.8: Calculated electron DOS for bulk Si using the modified Beck-Johnson potential.

Figure (5.9) shows our calculated band structure for Si. The bands are plotted along the symmetry points of the first Brillouin zone of the FCC lattice [12]. Even for the simple semiconductor, the curves are fairly complex. Notice that there exist an energy value for which all states below are occupied, and all states above are empty. This is the Fermi energy (E_F), and in Figure (5.9), the band structure is shifted so that E_F is at zero. Depending upon the position of E_F relative to the band structure, one can make a simple, but important, distinction concerning the electrical conductivity of the material:- If E_F cuts through the band, it is a metal. If it does not, it is a semiconductor (or insulator). Figure (5.9) says that Si is a semiconductor, and indeed it is. Notice that the region of energy just above E_F has no bands; this is the band gap. To put it briefly, the band structure is helpful to visualize the wavevector-dependence of the energy states, the band gap, and the possible electronic transitions.

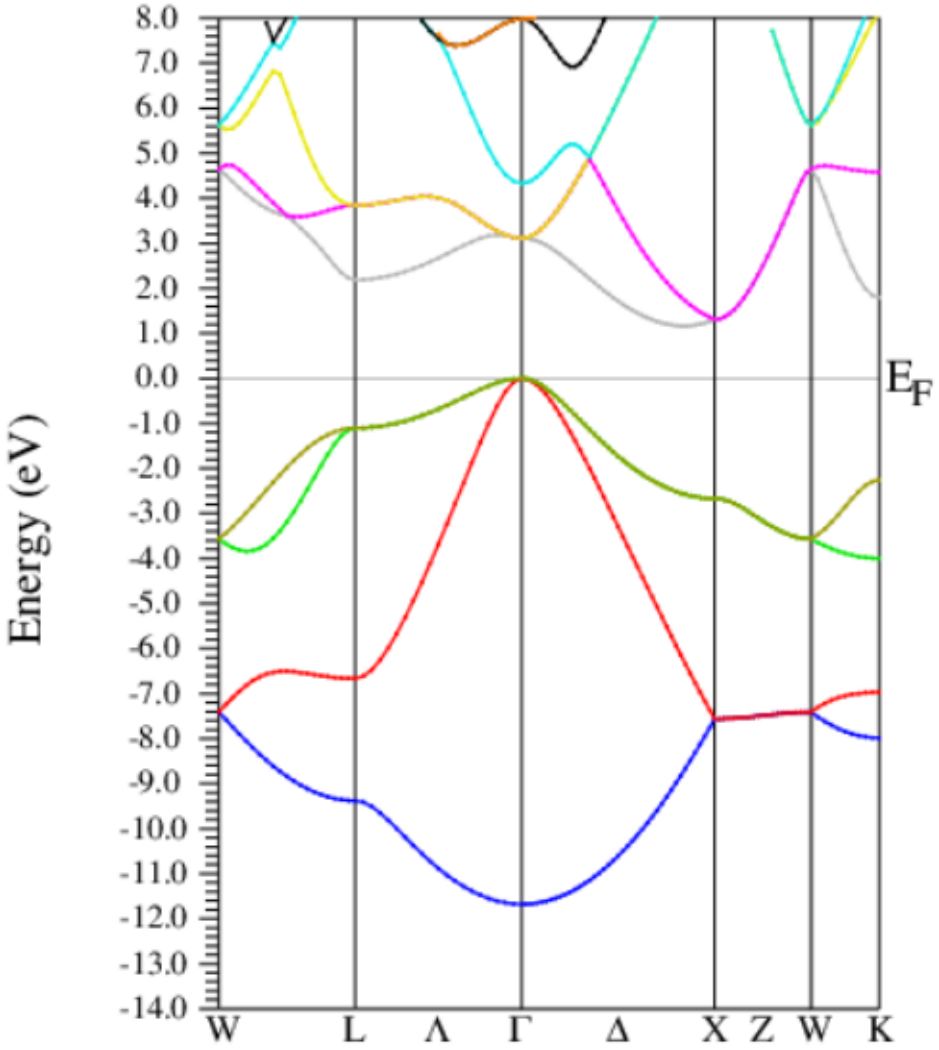


Figure 5.9: The energy bands for Si, calculated with the mBJ potential and 680 **k**-points in the IBZ. The bands are plotted along the high-symmetry points of the FCC lattice.

Chapter 6

First-Principles Study of Organic-Inorganic Perovskites

The efficient solar cell absorbs incident photons over a wide range of spectra, from visible to near-infrared wavelengths, and converts them into efficiently movable charges. The electronic structure of the efficient solar cell, then, should have a suitable band gap - one that allows the absorption of different photons in the spectrum. Hybrid organic-inorganic lead halide perovskites serve as the crown of the jewel in the field of emerging photovoltaic technology. Since the pioneering work by Kojima *et al.* [40] and Im *et al.* [41], the power conversion efficiency (PCE) has increased from a mere 3.8% to an outstanding 22.1% [50]. These materials contain a large and ever-surprising variety of properties (which include high carrier mobilities, low carrier recombination rates, and tunable spectral absorption range) and have the versatility to accommodate almost all of the elements of the periodic table. The remarkable range of structure and property interplay makes them an excellent research field in photovoltaic technology. Accordingly, there has been tremendous enthusiasm, interest, and growth from the scientific community; and in 2013, the editors of *Science* and *Nature* selected perovskite cell technology as one of the biggest scientific breakthroughs of the year [114][115].

6.1 Crystal Structures

Early studies reported that the perovskite structure, ABX_3, was primarily cubic. The A cation rests at the center of the cube, occupying the 12-fold

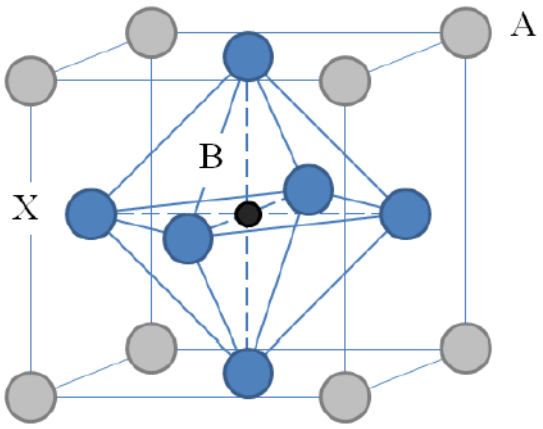


Figure 6.1: The ideal cubic perovskite structure of general form ABX_3. The akali atoms (Cs, Rb) occupy the A sites, Pb the B sites, and the halogen atoms (Cs, Br, I) the X sites [28].

coordination site shared with twelve X anions, while the B cation sits at the cube corners in a stabilized mesh of corner-sharing octahedra, as shown in Figure (6.1). As work continued, the number of proposed symmetries increased. Methylammonium lead iodide ($CH_3NH_3PbI_3$) and its derivatives are derived from ABX_3 (A=Cs, Rb; X=Cl, Br, I) by replacing the alkali-metal atom (A) with methylammonium (MA). They have established their place as one of the most exciting photovoltaic technologies. Unfortunately, perovskite cells have yet to find widespread adoption because of the lingering problems concerning the toxicity of lead and long-term stability. Fortunately, the electronic properties of these materials depend on various factors, such as lattice constants, types of halogen atoms, kinds of alkali metals, which can be controlled experimentally. Their fabrication methods are still relatively new, and so it is important to develop computational models which can accurately describe the known electronic properties so that the effects of variable structures can be predicted. The research question leading the current study is formulated accordingly: How do different atomic structures of the perovskite cell influence the band gap?

In this work, the electronic structure of five perovskite compounds, MA-$PbBr_3$, $CsPbX_3$ (X=Cl, Br, I) and $RbPbI_3$, are systematically studied from

Compound	Structure	Lattice Constant (Å)
$CsPbCl_3$	Cubic	$a = 5.605$[a]
$CsPbBr_3$	Orthorhombic	$a = 8.244$, $b = 11.735$, $c = 8.198$[b]
$CsPbI_3$	Orthorhombic	$a = 10.434$, $b = 4.790$, $c = 17.761$[c]
$MAPbBr_3$	Cubic	$a = 5.933$[d]
$CsPbI_3$ (cubic)	Cubic	$a = 6.2894$[c]
RbPbI3	Orthorhombic	$a = 10.276$, $b = 4.779$, $c = 17.393$[e]

[a] Nelmes *et al.* [116]

[b] Christos *et al.* [117]

[c] Mercouri *et al.* [118]

[d] Azetsu *et al.* [119]

[e] Myagkota *et al.* [120]

Table 6.1: The various crystal structures and corresponding lattice constants studied in this work.

first principles using the all-electron, full potential, linearized augmented plane wave ((L)APW) + local orbitals (lo) method as implemented in the WIEN2k code [73]. Their crystal structures and lattice constants are listed in Table (6.1) and fractional coordinates in Appendix C.

6.2 Calculation Methods

Here, space is divided into distinct regions: non-overlapping spheres about each atomic site and the remaining volume between them. This way, the Schrodinger equation can be solved exactly within each sphere, in atomic-like functions, and in the remaining interstitial region, in planewaves. The radii of the muffin-tin spheres R_{mt}, which are chosen to be as large as possible without touching, are 2.5 a_0 for Cs, Pb, I, Br, and Rb, 2.37 a_0 for Cl, 1.23 a_0 for C, 1.24 a_0 for N, and 0.67 a_0 for H, where a_0 is the Bohr radius. Inside the muffin-tin spheres, the basis functions, are expanded up to $l_{max} = 10$, while the planewaves in the interstitial region are expanded up to a wavevector cutoff K_{max} (which is customarily chosen such that $R_{mt}K_{max} = 6-9$). Given

the smallness of the muffin-tin radius of hydrogen atoms, $R_{mt}K_{max} = 3$ for MAPbBr3, while $R_{mt}K_{max} = 9$ for the remaining perovskites. Careful tests have been performed to check the convergence of the band gap results with the applied basis set and k-point mesh (and are shown in Appendix B).

The PBE exchange-correlation functional [93] has been used for all the calculations. The charge density is Fourier-expanded up to a maximum wavevector G_{max}, where $G_{max} = 20a_0^{-1}$ for MAPbBr3, and $G_{max} = 12a_0^{-1}$ for the remaining perovskites. In the self-consistent calculation, the total energy and charge were converged with a total tolerance of 0.1 mRy and 0.001e, respectively.

6.3 Results and Discussion

The calculated band structure of $MAPbBr_3$ along high symmetry points in the Brillouin zone (BZ) is shown in Figure (6.2). A $9 \times 9 \times 9$ Monkhorst-Pack grid [108] was used for sampling the BZ of the simple cubic cell. The valence band maximum (VBM) and conduction band minimum (CBM) occur at the symmetry point R(1/2, 1/2, 1/2) [12], as expected [121]. Illustrated in Figure (6.3) is the density of states; the low-lying conduction bands are composed of Pb states, while Br states dominate the valence bands. The calculated band gap, 1.675 eV, is summarized in Table (6.2).

The band structure of $CsPbBr_3$ is shown in Figure (6.4). Here, the $CH_3NH_3^+$ cation is exchanged for the more symmetrical Cs^+ cation, which modifies the structure the perovskite from cubic to orthorhombic. Calculations are made using the PBE functional, and a **k**-point mesh of $20 \times 14 \times 20$, suggest a gap of 1.834 eV at the BZ center, Γ(0,0,0). The density of states is shown in Figure (6.5), where it can be seen that the bands around the gap are composed of Pb and Br orbitals.

The systematical study of the chemical trends of the ABX_3 perovskite structures continues with the methodical exchange of the X(=Br) atom for the Cl and I atoms. On the one hand, the $CsPbCl_3$ compound has a cubic unit cell while that of the $CsPbI_3$ compound is orthorhombic. The band structure of $CsPbCl_3$ is shown in Figure (6.6). The self-consistent calculation was performed with a k-point mesh of $11 \times 11 \times 11$, and suggest a gap of 1.931 eV at the expected R(1/2,1/2,1/2) point in the BZ [121]. Figure (6.7) shows the density of states where it can be seen that the VBM is essentially composed of Pb states, and the CBM is derived from the Cl states.

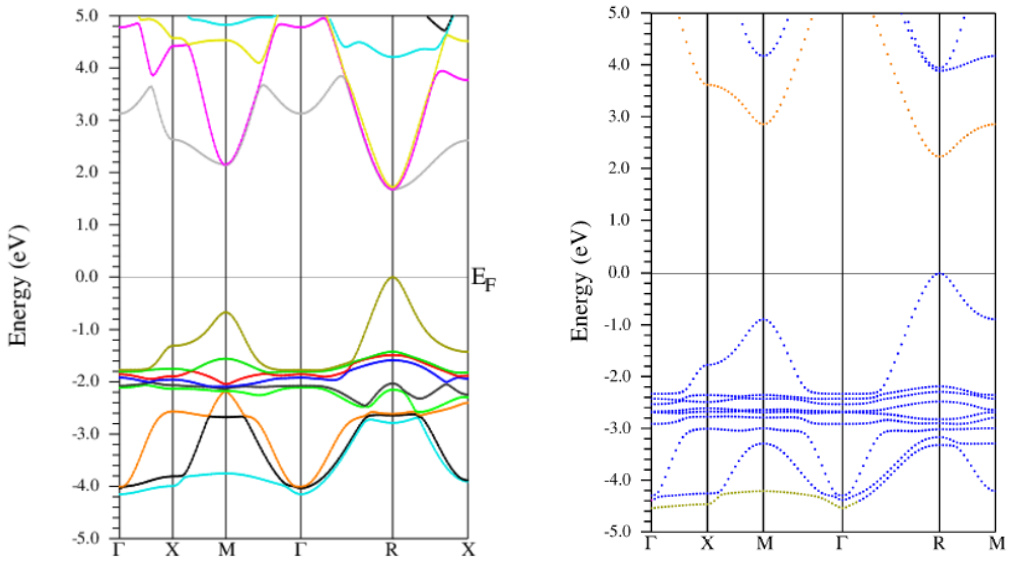


Figure 6.2: The band structure of $MAPbBr_3$: (a) our calculation; (b) from literature [121].

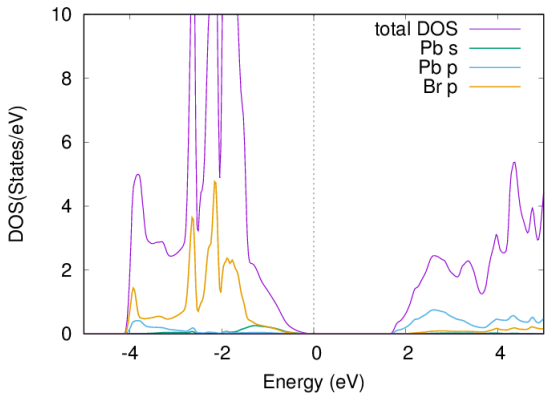


Figure 6.3: The density of states of $MAPbBr_3$

Compound	PBE	Experimental
$MAPbBr_3$	1.675	2.28[a]
$CsPbBr_3$	1.834	2.24[b]
$CsPbCl_3$	1.931	2.86[c]
$CsPbI_3$	2.512	3.14[d]
RbPbI3	2.454	3.17[d]
$CsPbI_3$ (cubic)	1.324	1.74[e]

[a] Heo *et al.* [122]

[b] Christos *et al.* [123]

[c] Stoumpos *et al.* [124]

[d] Kovalenko *et al.*[125]

[e] Hague *et al.* [126]

Table 6.2: The calculated and experimental band gaps, in eV, for the various perovskite structures studied in this work.

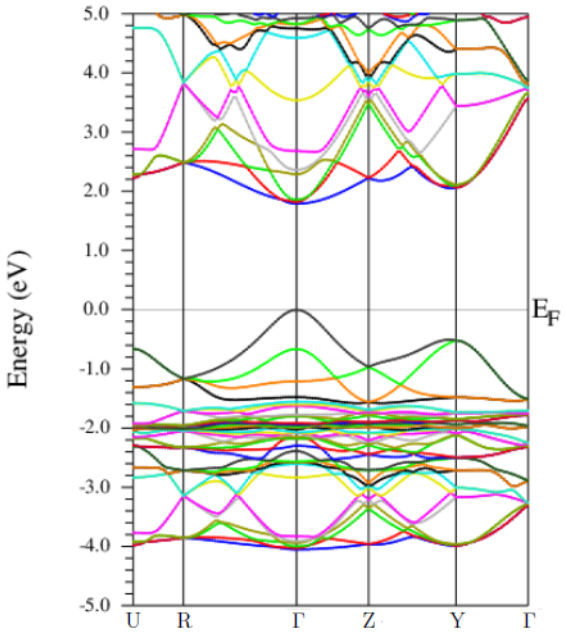
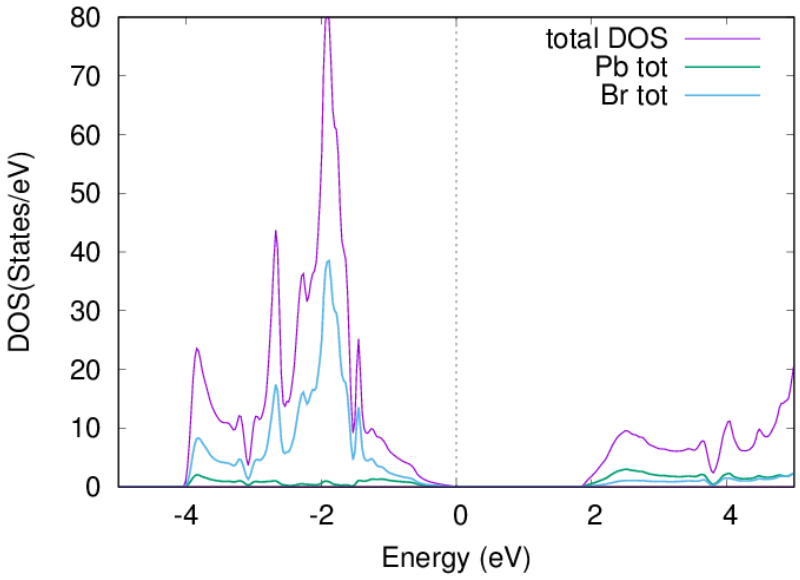
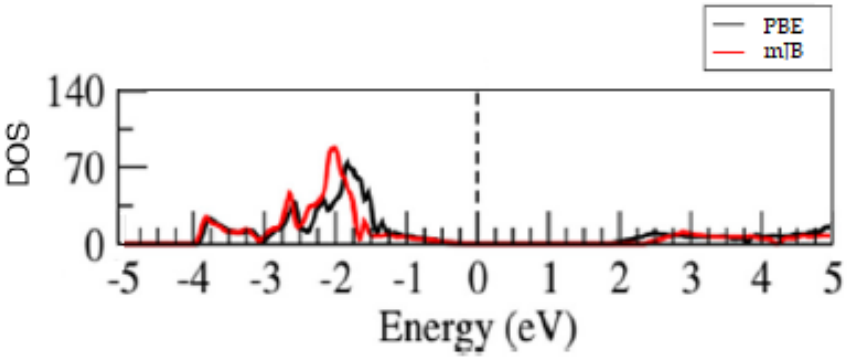


Figure 6.4: The band structure of $CsPbBr_3$



(a)



(b)

Figure 6.5: The band structure of cubic $CsPbBr_3$: (a) our calculation; (b) from literature [127].

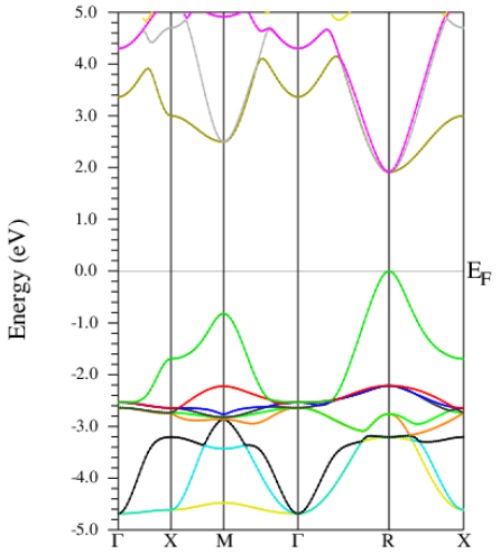
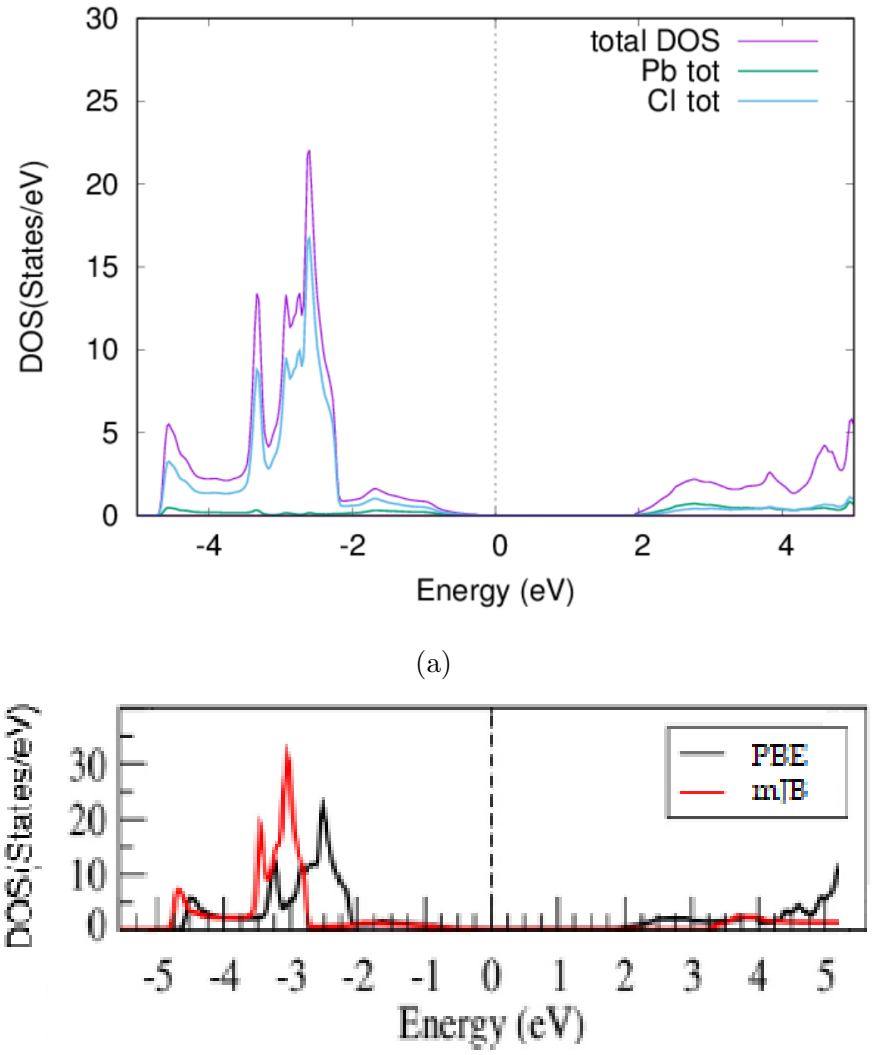


Figure 6.6: The band structure of $CsPbCl_3$.

The (room temperature) calculation for the orthorhombic $CsPbI_3$ crystal was performed with a k-point mesh of $10 \times 23 \times 6$ and suggest a band gap of 1.931 eV. As the temperature increases ($T > 634K$), the $CsPbI_3$ settles into a cubic structure and has a well-converged band gap of 1.324 eV. The corresponding band structure and density of states are shown in Figure (6.8) and Figure (6.9) respectively. There, it can be seen that the band gap occurs at the expected R point [121], and the bands around the gap are derived from Pb and I states.

Given the importance of lead-halide perovskites, a systematic and accurate investigation of the ABX_3 type compounds was accomplished to understand the nature of the electronic structures better. In this work, the band gaps of five room temperature perovskite compounds using first-principles calculations are systematically investigated. Their chemical trends as A and B vary are analyzed. It is noted that: (i) the band gap of ABX_3 increases when A changes from MA to Cs; (ii) as X changes from Br to Cl to I, the band gap increases; and (iii) as A changes from Cs to Rb, the band gap mostly remains the same.

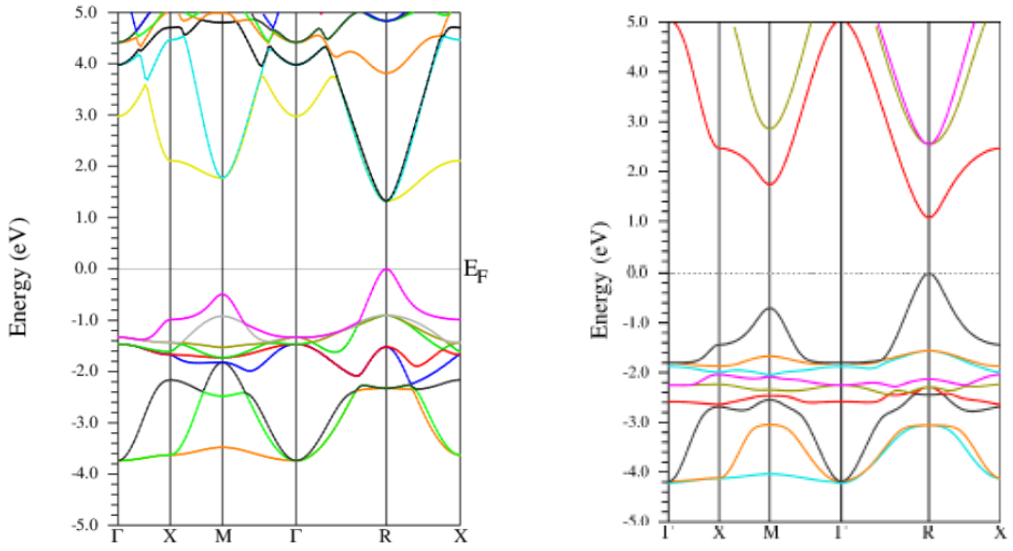
It is necessary to note that while the calculations correctly predict the chemical trends of these compounds, the widths of the gaps are not accurately predicted. This is, of course, a systematic error in the DFT calculations. It



(a)

(b)

Figure 6.7: The band structure of cubic $CsPbCl_3$: (a) our calculation; (b) from literature [127].



(a) (b)

Figure 6.8: The band structure of cubic $CsPbI_3$: (a) our calculation; (b) from literature [121].

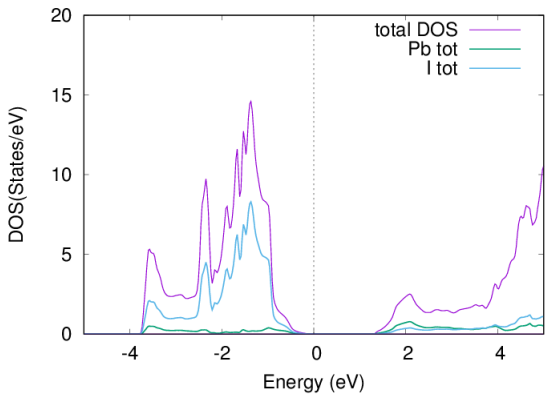


Figure 6.9: The density of states of cubic $CsPbI_3$.

is, therefore, important to explore different functionals that can, not only describe the chemical trends but also accurately describe their electronic structures [121] because how these materials work is still largely unknown.

Chapter 7

Summary and Outlook

The ill effects of climate change affect all trends, and the steps taken in the drive to reduce global emissions will reverberate for thousands of years. It is among the most significant and urgent problems to face, and so it is immensely important to call upon existing and near future technologies for generating clean electricity. For now, the most talked-about renewable energy source is solar. It is a massive resource by any standard and can easily provide the necessary energy services to accommodate the economic growth in energy consumption. It has the potential to play an essential role in decreasing the dependency on crude oil and reducing fossil fuel emissions. This work is motivated by the enormous potential and importance attached to photovoltaic devices to reduce global emissions and provide global energy for all purposes.

The technological base of photovoltaics is becoming progressively dependent on complicated materials. The fundamental piece of information needed to outline their characteristic properties is the wavefunction. Unfortunately, the enormous collection of interacting electrons present in these materials makes it impossible to solve for the wavefunction. The totality of every coordinate of every particle is mathematically fraught and so complicated that the scientist, who knows the arguments upon which the theory is founded, must appeal to physical plausibility. It, therefore, becomes beneficial to develop approximate practical methods which can lead to the explanation of the main features of the complex systems without too much computation.

Density functional theory (DFT) is a remarkably successful and versatile quantum mechanical approach to this problem. It is different than wavefunction based methods because it considers as the elementary variable the

electronic charge density, which is a function of only three coordinates. In this framework, the many-body Schrodinger equation is cast into a variational problem of minimizing an energy functional with respect to the electron density. However, given the energy functional (LDA, GGA, meta-GGA, etc.), the minimization problem can instead be resolved by finding a self-consistent solution to a set of single-particle equations - the Kohn-Sham equations. Fortunately, these equations can be solved iteratively as an eigenvalue problem with modest computing power. In other words, DFT can efficiently and accurately extract information from relatively large systems.

Much of the pioneering structural work on perovskites was done in the 1920s and formed the fundamental basis for further exploration of the perovskite family of compounds [28]. The corresponding structural class consists of a vast collection of compounds all having a crystal structure related to the mineral perovskite $CaTiO_3$ (ABX_3). These materials contain a large and ever-surprising variety of properties and have the versatility to accommodate almost all of the elements of the periodic table. Recently, methylammonium (MA) triiodideplumbate and its derivatives ($CH_3NH_3PbX_3$, X=I, Br, Cl) have attracted significant interest because of their potential application as light-harvesters. Today, the best-performing perovskite cells have reached a power conversion efficiency of 22.1% [50]. These high-efficiencies are mostly associated with the specific method of fabrication. The perovskite pigment, for instance, can be deposited in a single-step using a mixture of PbX_2 and CH_3NH_3X (X=Cl, Br, or I) from a common solvent (but generally leads to uncontrolled morphological variations and results in weak performances). The cell can also be fabricated in a sequential deposition method (which provides better control of the composition, thickness, and morphology of the crystal, and therefore, results in better performances). The large variety of fabrication techniques provides room for researchers to explore the largely debated fundamental physics behind the perovskite cell.

In the present work, the electronic structure of five perovskite compounds, $MAPbBr_3$, $CsPbX_3$ (X=Cl, Br, I) and $RbPbI_3$, are systematically studied from first principles using the all-electron, full potential, linearized augmented plane wave ((L)APW) + local orbitals (lo) method as implemented in the WIEN2k code. To assure the accuracy of the underlying ground state energies and band gaps, all calculations were carefully tested for convergence. It is noted that: (i) the band gap of ABX_3 increases when A changes from MA to Cs; (ii) as X changes from Br to Cl to I, the band gap increases; and (iii) as A changes from Cs to Rb, the band gap mostly remains the same.

The PBE functional employed in this work correctly predicts the chemical trends of the perovskite compounds but underestimates the band gaps by about 30%. This is a systematical error, and therefore, better results can be achieved with "better" functionals. It is important to use one that can not only describe the chemical trends but also accurately describe the electronic structure which may help to understand the origin of the photovoltaic performances of these materials.

It is well known that $CH_3NH_3PbI_3$ is unstable in humid conditions; it forms PbI_2, a carcinogen, at higher temperatures due to the loss of CH_3NH_3I [26]. Lead compounds are very toxic and damaging to the environment, and so these instabilities pose health risks and prevent outdoor applications. These issues must be discussed before a widespread application can be realized.

Appendix A

Variational Method

Mathematical modeling of many problems in physics involves partial differential equations (PDEs). The most frequently studied PDEs are of second order; the dynamics of many systems involve just two derivatives (∇^2)). In practical research, analytical solutions are typically not possible, and so one must rely on numerical solutions and techniques to solve them.

There are two standard approaches in quantum mechanics: the perturbation theory and the variational method. Perturbation theory is useful when there is a small dimensionless parameter in the problem such that the system can be solved exactly when it is sent to zero. This leads to an expression for the desired solution in a power series expansion in the small parameter. The variational method does not require that the nature of the system to have a small parameter (nor that the system is exactly solvable in a specific limit). Instead, the success of the method depends on the initial "guess" (ansatz) of the wavefunction. This method, therefore, relies on excellent physical intuition.

By definition, the ground state $|0\rangle$ has the lowest energy eigenvalue E_0 of a given system

$$H|0\rangle = E_0|0\rangle . \tag{A.1}$$

This means that any other states have higher energy eigenvalues,

$$H|n\rangle = E_n|0\rangle , \quad E_n > E_0. \tag{A.2}$$

With this, consider a Hamiltonian whose eigenstates and eigenvalues cannot be solved for exactly. The problem is to submit a guess of what the ground

state looks like based on sensible physical intuition; $\left|\tilde{0}\right\rangle$. If this ansatz is not explicitly correct, then it corresponds to a linear combination of *different* Hamiltonian eigenstates

$$\left|\tilde{0}\right\rangle = \sum_{n=0}^{\infty} |n\rangle \left\langle n\middle|\tilde{0}\right\rangle = \sum_{n=0}^{\infty} c_n |n\rangle . \tag{A.3}$$

Note that the ansatz is exact if $c_0 = 1$ and $c_n = 0$ for all $n \neq 0$, and the normalization condition requires

$$|c_0|^2 + \sum_{n\neq 0} |c_n|^2 = 1. \tag{A.4}$$

The energy expectation value computed with the ansatz is

$$\left\langle\tilde{0}\right| H \left|\tilde{0}\right\rangle = |c_0|^2 E_0 + \sum_{n\neq 0} |c_n|^2 E_n \geq |c_0|^2 E_0 + \sum_{n\neq 0} |c_0|^2 E_0 = E_0. \tag{A.5}$$

In other words, the expectation value must always be greater than or equal to the ground state energy [12]. One can, therefore, obtain an upper limit on the ground-state energy eigenvalue by employing the ansatz. Evidently, using a "better" trial ket produces a better overlap with the true ground-state - so that c_0 is closer to unity - and, by extension, a better upper limit is achieved. The converse is equally true: A state with a lower expectation value of the Hamiltonian should have a better overlap with the true ground-state. Therefrom results in an interactive scheme to find the true ground-state wavefunction:- Introduce parameters to the trial ket $\left|\tilde{0}(\lambda_1, \lambda_2, ...)\right\rangle$, calculate the energy expectation value,

$$\bar{E}(\lambda_1, \lambda_2, ...) = \left\langle\tilde{0}(\lambda_1, \lambda_2, ...)\right| H \left|\tilde{0}(\lambda_1, \lambda_2, ...)\right\rangle , \tag{A.6}$$

and look for as low an expectation value as possible,

$$\frac{\partial \bar{E}}{\partial \lambda_1} = \frac{\partial \bar{E}}{\partial \lambda_2} = \cdots = 0. \tag{A.7}$$

This is the variational method. It consists of choosing a trial wavefunction that depends on one or more parameters and finding the corresponding set so that the energy expectation value is the lowest possible. If the ansatz is close to the true ground-state, so too is the minimum expectation energy.

Appendix B

Convergence Tests

Compound	k-points (BZ)	No. of k Points (IBZ)	RKmax
$CsPbCl_3$	[11×11×11]	16	9.0
$CsPbBr_3$	[20×14×20]	600	9.0
$CsPbI_3$	[10×23×06]	180	9.0
$MAPbBr_3$	[09×09×09]	365	3.0
$CsPbI_3$ (cubic)	[11×11×11]	56	9.0
$RbPbI_3$	[10×23×06]	180	9.0

Table B.1: The employed Monkhorst-Pack grids and $R_{mt}K_{max}$ values during the self-consistent calculations.

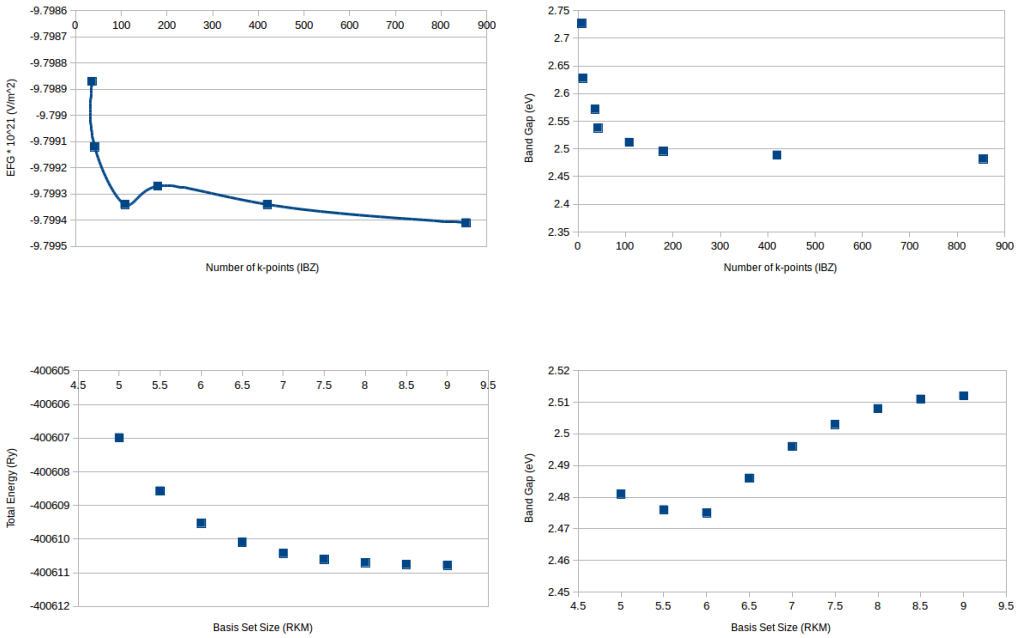


Figure B.1: The convergence tests for $CsPbI_3$. The electric field gradient and band gap are plotted as a function of k-points; the total energy and band gap are plotted as a function of basis set size.

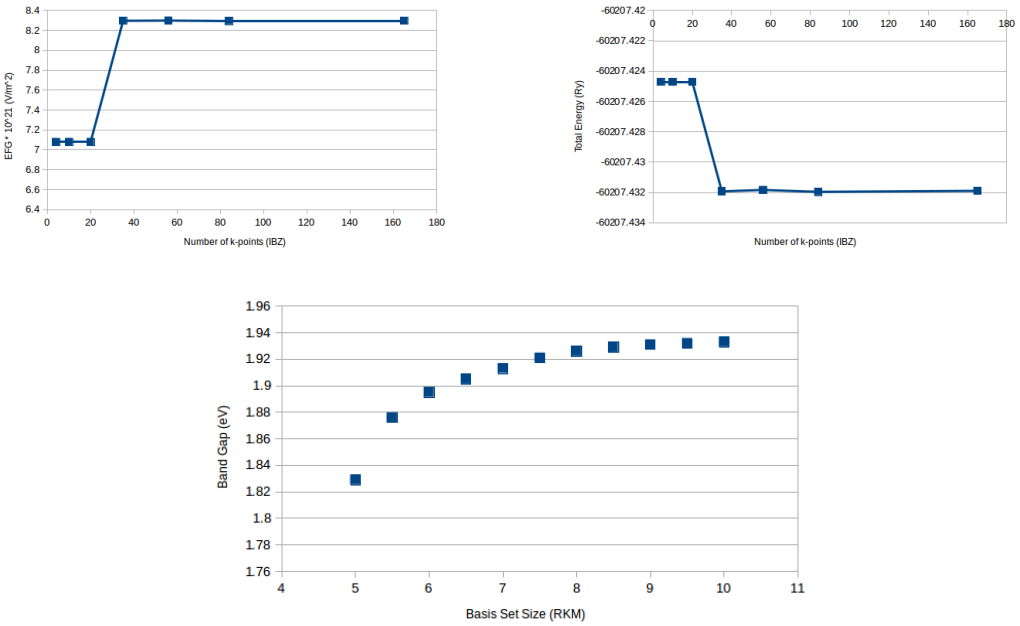


Figure B.2: The convergence tests for $CsPbCl_3$. The electric field gradient and total energy are plotted as a function of k-points; the band gap is plotted as a function of basis set size.

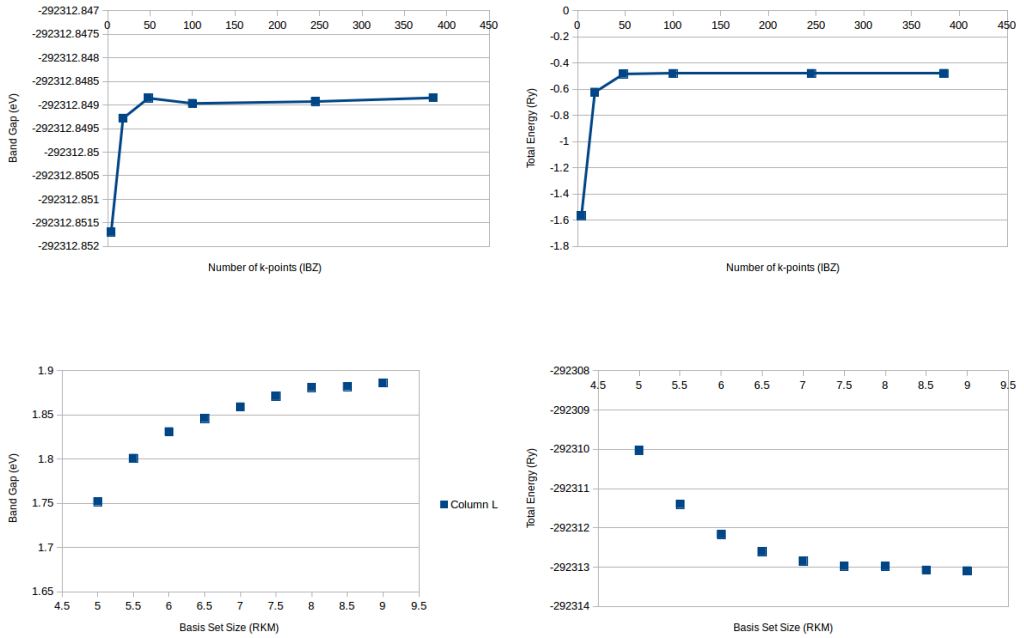


Figure B.3: The convergence tests for $CsPbBr_3$. The band gap and total energy are plotted as a function of k-points; the band gap and total energy are plotted as a function of basis set size.

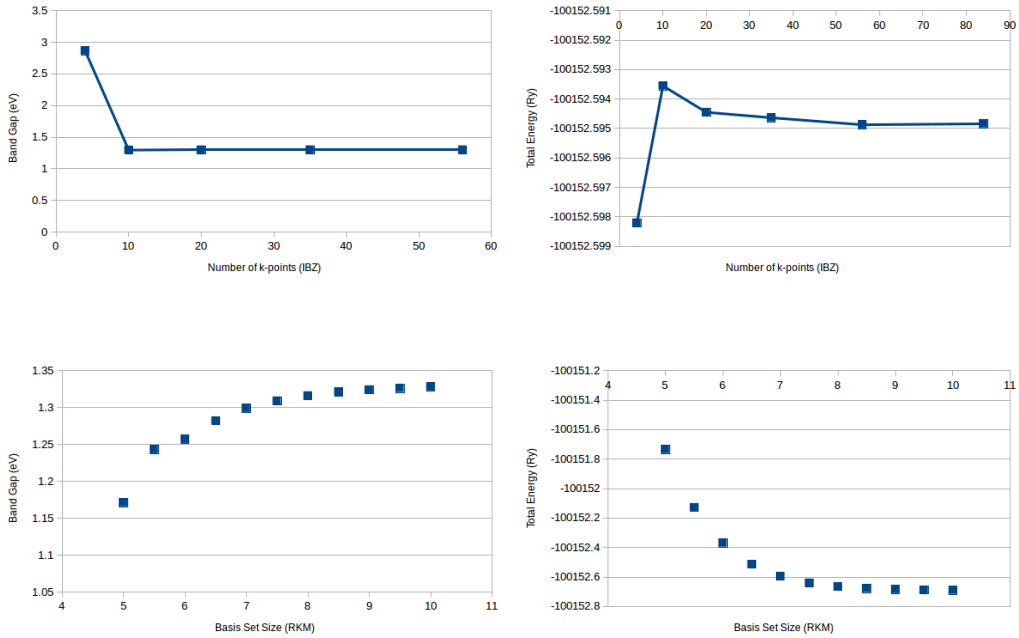


Figure B.4: The convergence tests for cubic $CsPbI_3$. The band gap and total energy are plotted as a function of k-points; the band gap and total energy are plotted as a funciton of basis set size.

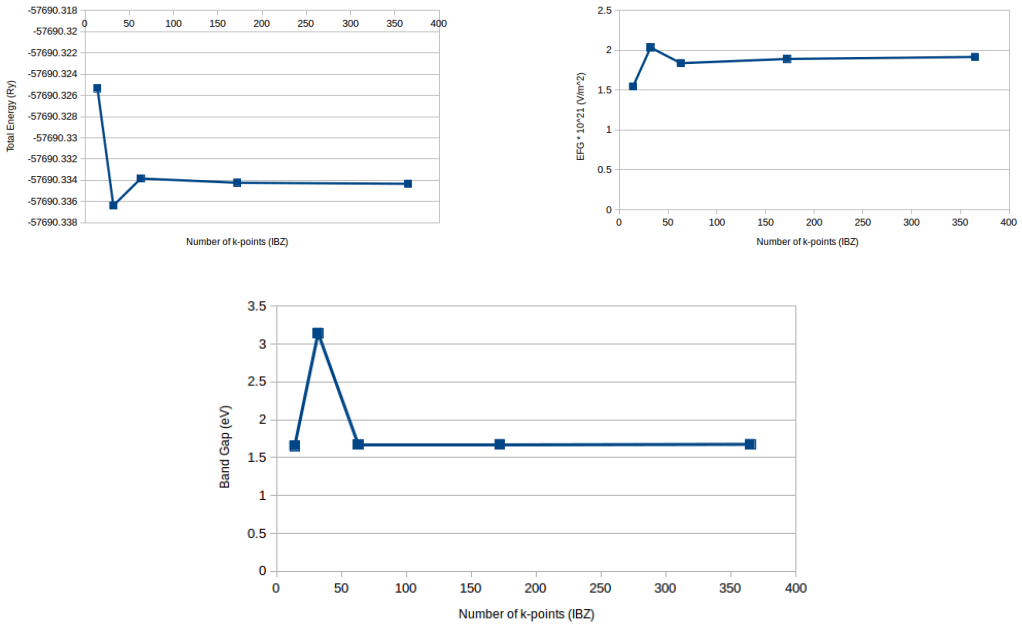


Figure B.5: The convergence tests for $CH_3NH_3PbBr_3$. The total energy and electric field gradient are plotted as a function of k-points; the band gap is plotted as a function of basis set size.

Appendix C

Crystal Structures

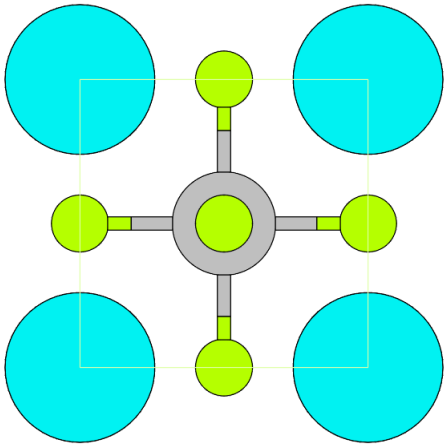
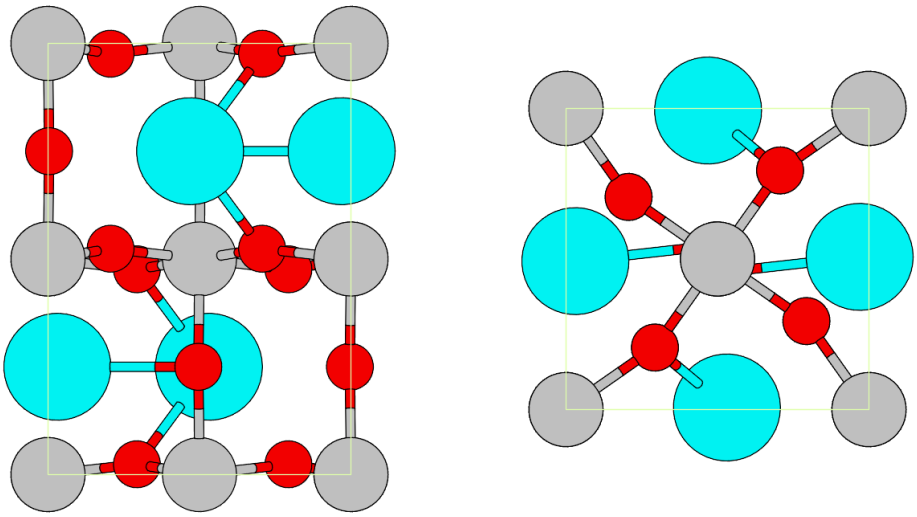
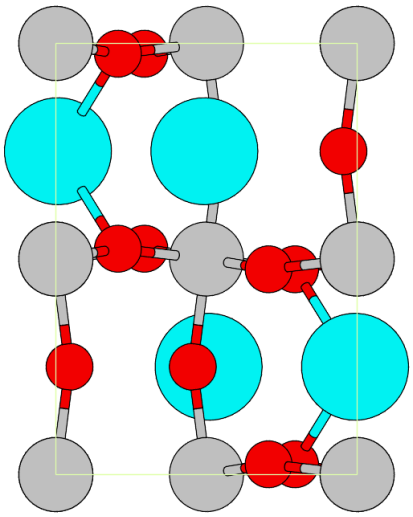


Figure C.1: $CsPbCl_3$ Crystal Structure [116] (cubic)



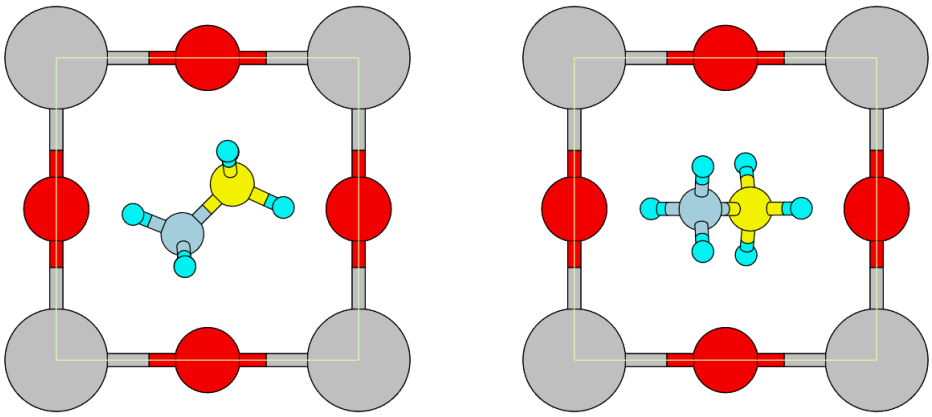
(a)

(b)



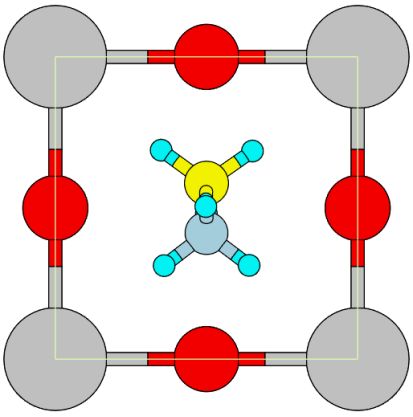
(c)

Figure C.2: $CsPbBr_3$ Crystal Structure [117]: (a) oriented to XY plane; (b) oriented to ZY plane; (c) oriented to YZ plane.



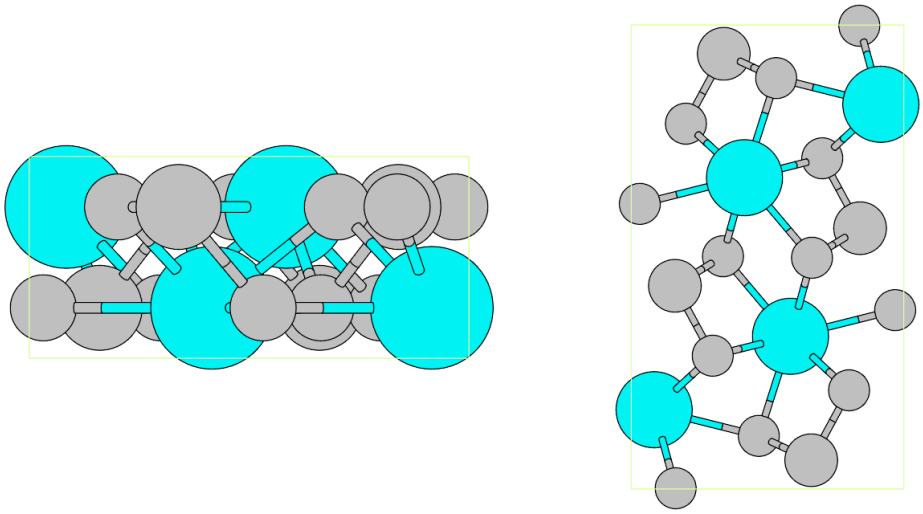
(a)

(b)



(c)

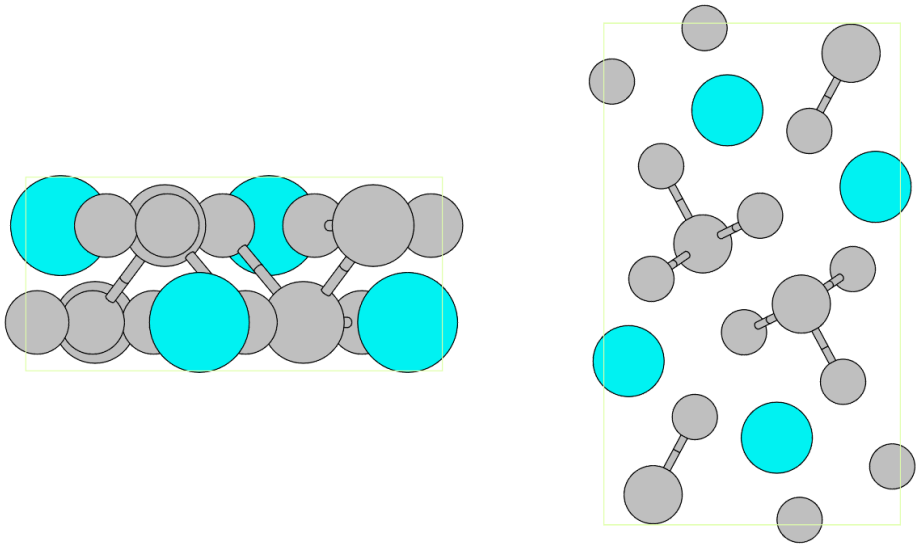
Figure C.3: $CH_3NH_3PbBr_3$ Crystal Structure [119]: (a) oriented to XY plane; (b) oriented to XZ plane; (c) oriented to YZ plane.



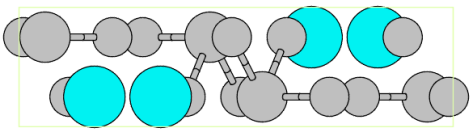
(a) (b)

(c)

Figure C.4: $CsPbI_3$ Crystal Structures [118]: (a) oriented to XY plane; (b) oriented to XZ plane; (c) oriented to YZ plane.



(a) (b)



(c)

Figure C.5: $RbPbI_3$ Crystal Structures [118]: (a) oriented to XY plane; (b) oriented to XZ plane; (c) oriented to YZ plane.

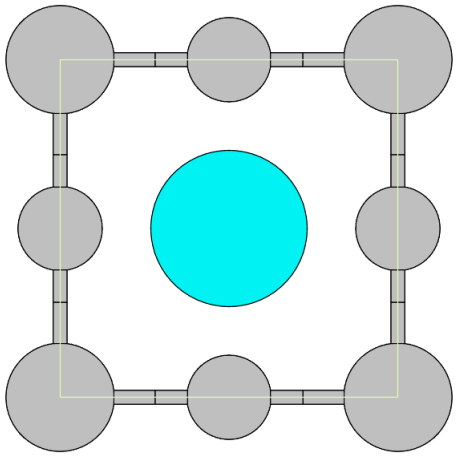


Figure C.6: $CsPbCl_3$ Crystal Structure [118] (cubic)

Space Group: 221 (Pm-3m)			
Lattice Vectors: $a = 10.5920\,\text{Å}$, $b = 10.5920\,\text{Å}$, $c = 10.5920\,\text{Å}$			
Fractional Coordinates:			
Pb	0.50000	0.50000	0.50000
Cl	0.00000	0.50000	0.50000
Cs	0.00000	0.00000	0.00000

Table C.1: Structural coordinates of $CsPbCl_3$ with experimental lattice constants.

Space Group: 62 (Pnma)			
Lattice Vectors: $a = 15.5790\,\text{Å}$, $b = 22.1761\,\text{Å}$, $c = 15.4924\,\text{Å}$			
Fractional Coordinates:			
Pb	0.50000	0.50000	0.00000
Cs	0.53100	0.25000	0.00702
Br	0.29347	0.47611	0.20607
Br	0.00370	0.75000	0.04550

Table C.2: Structural coordinates of $CsPbBr_3$ with experimental lattice constants.

Lattice Type: P			
Lattice Vectors: $a = 5.9330\,\text{Å}$, $b = 5.9330\,\text{Å}$, $c = 5.9330\,\text{Å}$			
Fractional Coordinates:			
Pb	0.00000	0.00000	0.00000
Br	0.00000	0.00000	0.50000
Br	0.50000	0.00000	0.00000
Br	0.00000	0.50000	0.00000
N	0.417000	0.41700	0.50000
H	0.253715	0.48209	0.50056
H	0.425296	0.30930	0.64000
H	0.425087	0.31017	0.35909
C	0.582000	0.58200	0.50000
H	0.751054	0.50500	0.50225
H	0.566350	0.69127	0.64990
H	0.568170	0.68848	0.34800

Table C.3: Structural coordinates of $CH_3NH_3PbBr_3$ with experimental lattice constants. The CH_3NH_3 positions are placed in appropriate starting positions compatible with the space group, and the crystal structure then is fully relaxed. The geometry of the crystallographic unit cell that minimizes the structures total energy is summarized here.

Space Group: 62 (Pnma)			
Lattice Vectors: $a = 10.4342\,\text{Å}$, $b = 4.790\,50\,\text{Å}$, $c = 17.7610\,\text{Å}$			
Fractional Coordinates:			
Pb	0.16039	0.25000	0.43797
I	0.16322	0.25000	0.00160
I	0.29910	0.25000	0.28729
I	0.03191	0.25000	0.61449
Cs	0.41564	0.25000	0.67090

Table C.4: Structural coordinates of $CsPbI_3$ with experimental lattice constants.

Space Group: 62 (Pnma)			
Lattice Vectors: $a = 10.276\,10$ Å, $b = 4.779\,32$ Å, $c = 17.393\,33$ Å			
Fractional Coordinates:			
Rb	0.41660	0.25000	0.82630
Pb	0.16620	0.25000	0.06001
I	0.30680	0.25000	0.21490
I	0.33960	0.75000	0.99020
I	0.02720	0.25000	0.88370

Table C.5: Structural coordinates of $RbPbI_3$ with experimental lattice constants.

Space Group: 221 (Pm-3m)			
Lattice Vectors: $a = 6.289\,42$ Å, $b = 6.289\,42$ Å, $c = 6.289\,42$ Å			
Fractional Coordinates:			
Cs	0.50000	0.50000	0.50000
Pb	0.00000	0.00000	0.00000
I	0.50000	0.00000	0.00000

Table C.6: Structural coordinates of cubic $CsPbI_3$ with experimental lattice constants.

Bibliography

[1] L. Benichou A. Roussel T. Clause Z. Kahraman C. Ringenbach. *The Shift Project Data Portal*. 2017. URL: http://www.tsp-data-portal.org/.

[2] "Statistical Review of World Energy". In: BP Global (2016). bp.com/statisticalreview.

[3] R.J. Andres T.A. Boden G. Marland. "Global, Regional, and National Fossil-Fuel CO2 Emissions. Carbon Dioxide Information Analysis Center". In: *Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, Tenn., U.S.A.* (2017). doi:10.3334/CDIAC/00001_V2017.

[4] G. Yohe C. Parmesan. "A globally coherent fingerprint of climate change impacts across natural systems". In: *Nature* 421 (2003). doi:10.1038/nature01286.

[5] National Petroleum Council (NPC). *Facing the Hard Truths About Energy*. 2007.

[6] M. A. Delucchi M. Z. Jacobson. "Providing all global energy with wind, water, and solar power, Part I: Technologies, energy resources, quantities and areas of infrastructure, and materials". In: *Energy Policy* 39 (2011).

[7] R. E. Dunlap A. M. McCright. "Defeating Kyoto: The Conservative Movement's Impact on U.S. Climate Change Policy". In: *Social Problems* 50 (2003).

[8] A. C. Kadak M. S. Kazimi E. J. Moniz J. E. Parsons J. M. Deutch C. W. Forsberg. "Update of the MIT 2003 Future of Nuclear Power". In: *Massachusetts Institute of Technology* (2009).

[9] International Energy Agency. *CO2 Emissions From Fuel Combustion Highlights 2016*. 2016.

[10] *Part 2: Solar Energy Reaching The Earth's Surface*. June 2018. URL: http://www.itacanet.org/the-sun-as-a-source-of-energy/part-2-solar-energy-reaching-the-earths-surface/.

[11] D. J. Mullan. *Physics of the Sun*. 2009.

[12] N. W. Ashcroft and N. D. Memin. *Solid state physics*. 1976.

[13] F. Geelhaar G. Heiser P. P. Altermatt A. Schenk. "Reassessment of the intrinsic carrier density in crystalline silicon in view of band-gap narrowing". In: *Applied Physics* 93 (2003).

[14] D. Tsamakis K. Misiakos. "Accurate measurements of the silicon intrinsic carrier density from 78 to 340 K". In: *Applied Physics* 74 (1993).

[15] S. Bowden C. Honsberg. *PVEducation*. June 2018. URL: https://www.pveducation.org/.

[16] J. M. Ziman. *Principles of the theory of solids*. 1965.

[17] D. Morel E. Stefanakos H. S. Ullal T. M.Razkov C. S. Ferekides. "Solar photovoltaic electricity: Current status and future prospects". In: *Sol. Energy* (2011).

[18] A. N. Hafez W. A. Badawy S. A. Elmeniawy. "Improvement of the power of industrially fabricated solar cells by etching of the Si surface and the use of surface analytical techniques". In: *Egypt. J. Anal. Chem.* 22 (2013).

[19] R. Knecht J. Parisi D. Heinemann W. Jurgens. "30 years at the service of renewable energies". In: *Research Journal of the University of Oldernburg, Germany* 54 (2011).

[20] M. Ohring. *Materials Science of Thin Films*. 2006.

[21] A. M. Hermann. "Polycrystalline thin-film solar cells - a review". In: *Sol. Energy* (1998).

[22] A. Khan M. Imaizumi S. Matsuda N. Daukes M. Yamaguchi T. Takamoto. In: *Photovoltaics* 26 (2018).

[23] J. Shi X. He L. Zhang D. Singh M. Lee S. Sie N. Suriyawong. In: *Photovoltaics* 26 (2018).

[24] B. Azzopardi. In: *Photovoltaics* 24 (2016).

[25] T. Satish Kumar N. Prabavathy S. Senthilarasu S. Prasanna S. Shalini R. Balasundaraprabhu. In: *Photovoltaics* 40 (2016).

[26] B. O'Regan and M. Gratzel. In: *Nature* 353 (1991).

[27] M. Gratzel. "The light and shade of perovskite solar cells". In: *Nature Materials* 13 (2014).

[28] R. Guo A. S. Bhalla and R. Roy. "The perovskite structure - a reivew of its role in ceramic science and technology". In: *Mat. Res* 4 (2000).

[29] A.P. Ramirez Y. Wang K.A. Regan N. Rogado M.A. Hayward M. K. Haas J.S. Slusky K. Inumaru1 H.W. Zandbergen1 N.P. Ong T. He Q. Huang and R.J. Cava. In: *Arxiv* (2001).

[30] C. Martin B. Raveau A. Maignan and M. Hervieu. In: *Chem. Mater.* 10 (1988).

[31] O. Takeo. *Crystal Structures of CH3Nh3PbI3 and Related Perovskite Compounds Used for Solar Cells*. 2015.

[32] H. Mashiyam and Y. Kurihara. "Disordered Cubic Perovskite Structure of CH3NH3PbX (X=Cl, Br, I". In: *Jornal of the Korean Physical Society* 32 (1998).

[33] J. M. Kadro M. Schreyer F. Wei S. G. Mhaisalkar M. Graetzel T. Baikie Y. Fang and T. J. White. "Synthesis and crystal chemistry of the hybrid perovskite (CH3NH3)PbI3 for solid-state sensitised solar cell applications". In: *J. Mater. Chem A1* (2013).

[34] H. Mashiyam Y. Kurihara and K. Hasebe. "Structural Study on Cubic-Tetragonal Transition of CH3NH3PbI3". In: *J. Phys. Soc. Jpn* 71 (2002).

[35] W. E. Guise K. Page Y. Q. Cheng I. Milas P.S. Whitfield N. Herron and M. K. Crawford. "Structures, Phases Transitions and Tricritical Behavior of the Hybrid Perovskite Methyl Ammonium Lead Iodid". In: *Scientific Reports* (2016).

[36] A. B. Walker F. Brivio and A. Walsh. "Structural and electronic properties of hybrid perovskites for high-efficiency thin-film photovoltaics from first-principles". In: *APL Mater* 1 (2013).

[37] D. S. Sholl and J. A. Steckel. *Density Functional Theory: A practical introduction*. 2009.

[38] S.Cottenier L.Marks and P. Blaha. *Optimization Notes*. 2004.

[39] M. K. Nazeeruddin S. Fantacci F. De Angelis. In: *J. Phys. Chem. C* 115 (2011).

[40] Y. Shirai T. Miyasaka A. Kojima K. Teshima. In: *J. Am. Chem. Soc.* 131 (2009).

[41] J. W. Lee S. W. Park J. H. Im C. R. Lee and N. G. Park. In: *Nanoscale* 3 (2011).

[42] J. H. Im K. B. Lee T. Moehl A. Marchioro S. J. Moon R. Humphry-Baker J. H. Yum J. E. Moser M. Gratzel H. S. Kim C. R. Lee and N. G. Park. In: *Sci. Rep.* 2 (2012).

[43] Z. Xue Q. Peng A. K. Chandiran B. Liu M. K. Nazeeruddin L. Etgar P. Gao and M. Gratzel. In: *J. Am. Chem. Soc.* 134 (2012).

[44] T. Miyasaka T. N. Murakami M. M. Lee J. Teuscher and H. J. Snaith. In: *Science* 338 (2012).

[45] S. J. Moon R. Humphry-Baker P. Gao M. K. Nazeeruddin J. Burschka N. Pellet and M. Gratzel. In: *Nature* 499 (2013).

[46] M. B. Johnston M. Liu and H. J. Snaith. In: *Nature* 501 (2013).

[47] T. Goto K. Uchida N. Miura M. Hirasawa T. Ishihara. In: *Phys. B. Condens. Matter* 201 (1994).

[48] R. C. R. Nagirir P. L. Burn P. Meredith Q. Lin A. Armin. In: *Nat. Photonics* 9 (2015).

[49] Y. Zhai K. Mielczarek W. Wang W. Ma A. Zakhidov Z. Vardeny C. Sheng C. Zhang. In: *Phys. Rev. Lett.* 114 (2015).

[50] E. Jung N. Jeon Y. Kim D. Lee S. Shin J. Seo E. Kim J. Noh S. Seok W. Yang B. Park. In: *Science* 356 (2017).

[51] *Research Cell Efficiency Records*. July 2018. URL: `https://www.energy.gov/eere/solar/downloads/research-cell-efficiency-records`.

[52] M. Born and R. Oppenheimer. "Zur Quantentheorie der Molekeln". In: *Ann. Phys. (Leipzig)* (1927).

[53] J. C. Slater. In: *Phys. Rev.* 35 (1930).

[54] U. Fano. In: *Phys. Rev.* 124 (1961).

[55] J. Bartlett and M. Musiał. In: *Rev. Mod. Phys.* 79 (2007).

[56] C. Møller and M. S. Plesset. In: *Phys. Rev.* 46 (1934).

[57] W. Kohn P. Hohenberg. In: *Phys. Rev.* 136 (1964).

[58] W. Kohn. "Electronic structure of matter - wave functions and density functionals". In: *Reviews of Modern Physics* 71 (1999).

[59] W. Kohn and L. J. Sham. "Self-Consistent Equations Including Exchange and Correlation Effects". In: *Phys. Rev.* 140 (1965).

[60] Viktor N. Staroverov. *Density-functional approximations for exchange and correlation*. 2013.

[61] L.H. Thomas. In: *Proc. Camb. Phil. Soc.* 23 (1927).

[62] E. Fermi Z. In: *Phys* 48 (1928).

[63] R. G. Parr and W. Yang. *Density-Functional Theory of Atoms and Molecules*. 1989.

[64] J.P. Perdew and W. Yang. In: *Phys. Rev.* 45 (1992).

[65] J.P. Perdew and S. Kurth. "Density functionals for non-relativistic Coulomb systems in the new century". In: *A Primer in Density Functional Theory* (2003).

[66] K. Burke J.P. Perdew and Y. Wang. "Generalized gradient approximation for the exchange-correlation hole of a many-electron system". In: *Phys. Rev.* 54 (1996).

[67] K. Burke J.P. Perdew and M. Ernzerhof. "Generalized Gradient Approximation Made Simple". In: *Phys. Rev.* 77 (1996).

[68] J. Tao V.N. Staroverov G.E. Scuseria G.I Csonka J.P. Perdew A. Ruzsinszky. "Prescription for the design and selection of density functional approximations: More constraint satisfaction with fewer fits". In: *The Journal of Chemical Physics* 123 (2005).

[69] C. Herring. In: *Phys. Rev.* 57 (1940).

[70] D. Singh. *Planewaves, Pseudopotentials and the LAPW method*. 2006.

[71] O.K. Andersen. In: *Phys. Rev. B* 12 (1975).

[72] A.J. Freedman H. Krakauer M. Posternak. In: *Phys. Rev. B* 19 (1979).

[73] G.K.H. Madsen D.Kavasnicka J. Luitz P. Blaha K. Schwarz. "WIEN2k, An Augmented Plane Wave Plus Local Orbitals Program for Calculating Crystal Properties". In: (2001).

[74] D. Singh and H. Krakauer. In: *Phys. Rev. B* 43 (1991).

[75] D. Singh and J. Ashkenazi. In: *Phys. Rev. B* 46 (1992).

[76] L. Nordstrom E. Sjostedt and D.J. Singh. In: *Solid State Commun.* 114 (2000).

[77] J. Schweifer K. Schwarz P. Blaha. *From crystal structure to properties of solids with the grid-enabled WIEN2K*. 2006.

[78] M. Weinert. "Solution of Poisson's equation: Beyond Ewald type methods". In: *Journal of Mathematical Physics* 22 (1981).

[79] D. D. Koelling and G. O. Arbman. In: *Phys. F: Met. Phys.* 8 (1975).

[80] L. D. Marks and R. Luke. In: *Phys. Rev. B* 141 (2008).

[81] L. D. Marks. In: *J. Chem. Theory Comput.* 141 (2013).

[82] G. Van Oost K. Lejaeghere V. Van Speybroeck and S. Cottenier. In: *Critical Reviews in Solid State and Materials Sciences* 39 (2014).

[83] Donald E. Sands. *Introduction to Crystallography*. 2013.

[84] Mariusz Jaskolski Zbigniew Dauter. "How to read (and understand) Volume A of International Tables for Crystallography: an introduction for nonspecialists". In: *J. Appl. Cryst* 43 (2010).

[85] A. L. Patterson. In: *Phys. Rev.* 46 (1934).

[86] A. McCoyb P. Evansa. In: *Acta Crystallogr D Biol Crystallogr* 64 (2008).

[87] B. Strandberg D. Davies D. Phillips V. Shore J. Kendrew R. Dickerson. In: *Nature* 185 (1960).

[88] P. Hedstrom R. Pelaez. In: *Critical Reviews in Solid State and Materials Sciences* 43 (2018).

[89] M. Narbe and H. G. Martin. "Thirty years of density functional theory in computational chemistry: an overview and extensive assessment of 200 density functionals". In: *Molecular Physics* 115:19 (2017).

[90] J. P. Perdew and K. Schmidt. "Jacob's ladder of density functional approximations for the exchange-correlation energy". In: American Institute of Physics Conference Series 577 (2001).

[91] I. N. Yakovkin and P. A. Dowben. "The problem of the band gap in LDA calculations". In: *Surface Review and Letters* 14 (2007).

[92] J. P. Perdew K. Burke and Y. Wang. "Electronic Density Functional Theory: Recent Progress and New Directions". In: *Plenum Press, New York* (1997).

[93] M. Ernzerhof J. P. Perdew K. Burke. In: *Phys. Rev. Lett.* 77 (1996).

[94] V. N. Staroverov J. Tao J. P. Perdew and G. E. Scuseria. In: *Phys. Rev. Lett.* 91 (2003).

[95] M. Ernzerhof K. Burke J. Perdew. *Electronic Density Functional Theory: Recent Progress and New Directions*. 1998.

[96] A. D. Becke. In: *J. Chem. Phys.* 98 (1993).

[97] A. D. Becke. In: *Phys. Rev. A* 38 (1988).

[98] R. Parr C. Lee W. Yang. In: *Phys. Rev. B* 37 (1988).

[99] J. P. Perdew and M. Levy. In: *Phys. Rev. Lett.* 51 (1983).

[100] P. Mori-Sanchez A. J. Cohen and W. Yang. In: *Chem. Rev.* 112 (2012).

[101] M. Levy. In: *Phys. Rev. A* 26 (1982).

[102] F. Tran and P. Blaha. In: *Phys. Rev. Lett.* 102 (2009).

[103] A. D. Becke and E. R. Johnson. "A Simple Effective Potential For Exchange". In: *J. Chem.* 124 (2006).

[104] A. D. Becke and M. R. Roussel. "Exchange Holes in Inhomogeneous Systems: A Coordinate-Space Model". In: *Phys. Rev. A* 39 (1989).

[105] F. Tran D. Koller and P. Blaha. "Merits and Limits of the Modified Becke-Johnson Exchange Potential". In: *Phys. Rev. B* 83 (2011).

[106] J. A. Camargo-Mertinez and R. Baquero. "The band gap problem: the accuracy of the Wien2k code confronted". In: *Revista Mexicanna de Fisica* 59 (2013).

[107] R. W. Hamming. *Numerical Methods for Scientist and Engineers*. 1987.

[108] H. J. Monkhorst and J. D. Pack. "Special points for Brillouin-zone integrations". In: *Physical Review B* 13 (1976).

[109] D. J. Chadi and M. L. Cohen. In: *Physical Review B* 8 (1973).

[110] S. L. Cunningham. In: *Physical Review B* 10 (1974).

[111] C. Kittel. *Introduction to Solid State Physics*. 2005.

[112] S. Cottenier. *Density Functional Theory and the family of (L)APW-methods: a step-by-step introduction*. 2013.

[113] *Advanced Solid State Physics*. June 2018. URL: `http://lamp.tu-graz.ac.at/~hadley/ss2/fermigas/dft/si/Si.php`.

[114] Science News. In: *Science* 342 (2013).

[115] In: *Nature News Features* 504 (2013).

[116] R. J. Nelmes J. Hutton. In: *Journal of Physics C* 14 (1981).

[117] D. Christos A. J. Peters Z. Liu M. Sebastian J. Im C. T. Chasapis C. A. Wibowow Y. D. Chung J. A. Freeman W. B. Wessels G. M. Kanatzidis C. C. Stoumpos D. C. Malliakas. In: *Crystal Growth and Design* 13 (2013).

[118] K. G. Mercouri S. C. Constantinos M. D. Christos. In: *Inorganic Chemistry* 52 (2013).

[119] T. Azetsu H. Mashiyam Y. Kurihara. In: *J. Korean Physical Soc.* 32 (1998).

[120] V. Myagkota M. Trots. In: *J. Phys. Chem. Solids* 69 (2008).

[121] A. Sharif R. Jishi O. Ta. In: *arXiv* 1405 (2014).

[122] H. Heo N. Mandal I. Seok H. Noh H. Im. In: *Nano Lett.* 13 (2013).

[123] A. Peters Z. liu M. Sebastian J. Im C. Chasapis C. Wibowo Y. Chung J. Freeman W. Wessels G. Kanatzidis C. Stoumpos D. Malliakas. In: *Cryst. Growth Des.* 13 (2013).

[124] C. Stoumpos M. Sebastian W. Wessels J. Im A. Freeman G. Kanatzidis Z. Liu J. Peters. In: *Proc. SPIE* 8852 (2013).

[125] N. Kovalenko N. Yunakova K. Miloslavaskii. In: *Functional Materials* 20 (2013).

[126] A. Hague E. Alarousu P. Sarmah B. Murali I. Dursun H. Miao L. Abdelady T. Wu F. Mohammed A. Zhumekenov I. Saidaminov. In: *ACS Energy Letters* 1 (2016).

[127] L. Ali M. Shafig R. Iqbal R. Ahmad T. Khan J. Asadabadi M. Magbool I. Ahmad M. Ahmad G. Rehman. In: *Journal of Alloys and Compounds* 705 (2017).